5-2012

Identifying the Spatial Distribution of Three Plethodontid Salamanders in Great Smoky Mountains National Park Using Two Habitat Modeling Methods

Matthew Stephen Kookogey
mkookoge@utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Matthew Stephen Kookogey entitled "Identifying the Spatial Distribution of Three Plethodontid Salamanders in Great Smoky Mountains National Park Using Two Habitat Modeling Methods." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Geography.

Liem Tran, Major Professor

We have read this thesis and recommend its acceptance:

Henri D. Grissino-Mayer, Frank T. van Manen

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Identifying the Spatial Distribution of Three Plethodontid Salamanders in Great Smoky Mountains National Park Using Two Habitat Modeling Methods

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Matthew Stephen Kookogey
May 2012

ACKNOWLEDGEMENTS

I want to thank my advisor, Dr. Liem Tran, for his continuous help on this project over the past few years. He was readily available when I had questions, and constantly pushed me to think appropriately about my topics. Often, I was frustrated or completely stumped but his guiding hand frequently led me in the right direction while simultaneously forcing me to find my own solutions. I have grown tremendously.

I also want to thank my committee members for their service in this project. Dr. Henri Grissino-Mayer has been a source of encouragement inside and outside matters of my thesis. His enthusiasm during meetings has helped boost my confidence from the first day on campus. Dr. Frank van Manen has also been a blessing, and without his class (and his employment!), I would still have a long way to go. The knowledge I gleaned from him came at critical junctures during this process. I would like to thank Keith Langdon of Great Smoky Mountains National Park for happily providing a taxon with which to work. I give many thanks to Dr. Thomas Chapman for introducing me to GIS with perfect timing and to Dan Coleman who cemented my interest in the natural world.

ABSTRACT

The main objective was to create habitat models of three plethodontid salamander species (*Desmognathus conanti*, *D. ocoee*, and *Plethodon jordani*) in GSMNP. To investigate the relationships between salamanders and their habitats, I used three models—logistic regression with use-availability sampling, logistic regression with case-control sampling, and Mahalanobis distance (D^2)—for each species to gain a robust view of the relationships. The secondary objective was to compare the different modeling methods within and across the three species. Elevation was the dominant variable for all three species.

D^2 for *D. conanti* predicted low elevations, close proximity to streams, metasandstones, and previously disturbed areas. The use-availability model indicated habitat in low elevation, settled areas, pine understory, and flood overstory. The case-control model for included only elevation and undisturbed areas. Case-control and D^2 predicted presences >90% correctly but absences <50% correctly. Use-availability was more balanced with 75% presences and 60.5% absences correct.

D. ocoee occurred only at the highest elevations. D^2 was influenced by proximity to streams, undisturbed areas, northern hardwood overstory, and frigid sandstones. Use-availability included a positive association with increasing elevation and a negative association with spruce understory. Case-control included elevation only. Use-availability did poorly at predicting presence (48.7%). Case-control predicted presences and absences at 79.5% and 89.0%, respectively. D^2 classified presences and absences at 92.3% and 78.1%, respectively.

P. jordani was also determined by elevation, but lower than *D. ocoee*. D^2 also included proximity to streams, undisturbed areas, and sandstones. Use-availability had a

negative association with spruce understory. Pine understory, northern hardwood overstory, and distance to streams were negatively associated with *P. jordani* occurrence in case-control. D^2 predicted 97.1% of presences and 66.0% of absences correctly. Use-availability predicted 64.1% of presences and 91.8% of absences correctly. Case-control predicted 74.8% of presences and 89.1% of absences correctly.

Use-availability worked best for *D. conanti*, but may have been a result of the uncertain identification of the species. For the other two species, case-control had a high classification rate for both presences and absences and a more intuitive answer for what determines habitat than D^2.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1 INTRODUCTION

1.1 BACKGROUND AND JUSTIFICATION

Land managers are always searching for effective ways to locate and protect species on their land. Because conservation, particularly of species not federally listed, is poorly funded, field work must be efficient if conservation is to be effective. Understanding where targeted species exist, their abundance, and the habitat characteristics they prefer is essential for conservation (Kiester et al 1996; Funk and Richardson 2002). Regional definitions of habitat can help describe an overall distribution of a species, but fieldwork based on a resolution too coarse may make fieldwork inefficient (Guisan et al. 2006). Land managers may need more specific clues on the habitat relevant to their jurisdiction. However, fine-scale definitions of resource use can describe the daily movements of an individual but cannot be extrapolated beyond the observations. Whereas habitat can be described differently, yet correctly, at different scales (Levin 1992), appropriately scaled models can help managers locate the best habitat for a species, therefore aiding conservation efforts, such as species reintroductions (Scott et al. 1993; Johnson and Gillingham 2005).

GSMNP is an area with high species diversity and abundant field observations exist based on the recently completed All Taxa Biodiversity Inventory (ATBI). The ATBI was a coordinated effort among scientists and citizen volunteers to locate and identify every species within GSMNP. The park is home to numerous species of plants and animals but is one of the most diverse areas in the world for salamanders. With over 30 species,

salamander diversity in GSMNP is comparable to tropical areas (Kozak and Weins 2007). Understanding salamander distributions is important because amphibians seem vulnerable to changes in environmental conditions and may be harbingers of future environmental conditions (Vitt et al. 1990; Welsh and Droege 2001; Davic and Welsh 2004). Upon meeting with Keith Langdon, the Inventory and Monitoring Coordinator of the GSMNP, I learned that very little is known about the distribution of many salamander species and their resource use.

My thesis focused on three salamander species in GSMNP and their distributions within the park boundaries. The distributions were determined using a presence-only method (Mahalanobis distance) and two variations of a presence-absence method (logistic regression), which I compared and contrasted. The salamanders I focused on were *Desmognathus conanti*, *D. ocoee*, and *Plethodon jordani*—all terrestrial salamander species that belong to the family Plethodontidae. The results from this study could help GSMNP understand the habitat requirements for these salamander species and provide direction for future biogeographic research.

1.2 Plethodontids and Existing Research

1.2.1 Family Plethodontidae

Salamanders from the family Plethodontidae are an exceptionally relevant group to study, comprising 25 of the salamander species in GSMNP. Furthermore, plethodontids are significant among salamanders because they are both lungless and terrestrial. Many species do not require flowing or standing water at any life stage (Petranka 1998). The family is ubiquitous and species exist in a variety of habitats in GSMNP (see Dodd 2004),

perhaps stemming from their evolutionary history. That history is complicated, but the current hypothesis is they lost their lungs in the late Cretaceous when the Appalachians were part of the Coastal Plain (Means 2000). Some speciation in the Appalachians may have happened with the uprising of the mountain chain evident in common habitat and resource use between related species on different mountain tops (Kozak and Wiens 2006). Other plethodontids differ in their life history and use of habitat, such as those that have aquatic larvae and therefore a greater dependency on standing and flowing water (Petranka 1998). I focused on three of the many plethodontid salamanders in GSMNP that exhibited distinct differences in their use of space.

1.2.2 Red-cheeked Salamander (*Plethodon jordani* Blatchley)

Plethodon jordani, the red-cheeked salamander, is a fairly abundant species and is endemic to GSMNP (Dodd 2004). The species is terrestrial and occupies areas of cool, humid forest floor, most easily observed in the summer after rainfall (Petranka 1998). Inedible to most would-be predators, individuals in the Smoky Mountains are distinguished by reddish cheeks, and are mimicked by the edible *Desmognathus imitator*. Most populations are found from 600 to 925 m elevation (Hairston 1949), although Dodd (2004) also found them at Clingmans Dome, the highest peak in GSMNP. Individuals generally have a small home range, only recorded to venture up to 11 m away from initial capture sites (Madison 1969). The observations in the dataset for this study had an elevation range of 537–2,016 m and indicated a preference for previously undisturbed forest.

1.2.3 Ocoee Salamander (*Desmognathus ocoee* Nicholls)

Desmognathus ocoee, the Ocoee salamander, is a terrestrial species usually found on wet rock faces and near streams. Populations at higher elevations tend to be more terrestrial than the populations at lower elevations (Petranka 1998). *D. ocoee* also has a moderately restricted Appalachian range, occurring mainly within GSMNP, Nantahala National Forest (North Carolina), and Chattahoochee National Forest (Georgia). A small, separate group also exists in the Alabama foothills of the Appalachians. Petranka (1998) stated this species has a larger elevation range than any other *Desmognathus* species, but in GSMNP it is restricted to the highest elevations of the park (Dodd 2004). Overall, not much is known about the Ocoee salamander's habitat requirements. The observations used for this study indicated a narrow range of elevation from 1,463 to 1853 m, and a preference for undisturbed, northern hardwood forests.

1.2.4 Spotted Dusky Salamander (*Desmognathus conanti* Rossman)

Desmognathus conanti, the spotted dusky salamander, is the most widespread of the three species. The salamander is found from southern Illinois to the Gulf of Mexico, west to central Louisiana and east to central South Carolina, staying north of the Coastal Plain of Georgia. Predominantly streamside species, *D. conanti* is generally found in wooded areas around trickling water (Petranka 1998). The species is more common at lower elevations but the range from the Gulf to Appalachia suggests a tolerance for conditions that coincide with increasing elevation. Observations from Dodd (2004) showed a high range of elevation (336–1,566 m), but the salamander was uncommon at elevations higher than

1,200 m. Also, the species was overwhelmingly found in areas that were settled prior to the establishment of GSMNP.

1.2.5 Existing Research

Existing research on salamander habitat fits into two categories. The first category includes detailed descriptive studies in which variables such as coarse woody debris and soil pH were measured (Ford et al. 2002; Crawford and Semlitsch 2008*a*; Crawford and Semlitsch 2008*b*). No landscape-scale data exist for these types of variables as they are very site specific and therefore cannot be integrated into broader landscape-scale habitat models. The second category includes habitat models that extrapolated results at a very broad scale, almost exclusively at a 1 km resolution. For example, Kozak and Weins (2006) determined modes of speciation of several salamanders from the family Plethodontidae in the Appalachian Mountains by examining range overlap. At such a coarse resolution landscape features are generalized, and using the model for detection and management of a species may be inefficient (Guisan et al. 2006).

Despite studies indicating that macro-variables are significant predictors of amphibian habitat (Diller and Wallace 1996; Russell et al. 2004; Stoddard and Hayes 2005), existing models are coarse. For conservation purposes, not only is 1 km too large for identifying habitat of a small animal such as a salamander, but in some habitats (i.e. mountainous), variables used as parameters in the model may change significantly over a 1 km^2 cell. Improvements in remote sensing data and landscape classification provide more precise data with high resolutions (e.g., 30 x 30 m). Specifically for GSMNP, no fine-scaled

salamander habitat models exist that would be beneficial for focused conservation purposes or improved detection in the field.

1.3 Habitat Models

1.3.1 Overview

Researchers have attempted to accurately predict species habitat and range using varying modeling techniques for years (Hairston 1949; Whittaker 1967; Gauch and Chase 1974). Habitat analysis shifted from largely qualitative descriptions of habitat to more quantitative evaluations of habitat gradients in the mid-1900s, but was ultimately limited computationally until the advent of personal computers (Stauffer 2002). Now with the ability to analyze variables simultaneously in a grid-based format, models provide a much more accurate determination of where favorable habitat exists and quantitative descriptions of covariates limiting habitat (Scott et al. 1993).

Data-based habitat models generally fit into two categories, presence-absence models and presence-only models. Most data best fit into presence-only models because the data were collected for studies only concerned with where the target species existed (Rotenberry et al. 2002). The nature of data collection is that field notes and coordinates are generally limited to where a species was observed. Even with attempts to record absence, determining true absence is challenging with rare species (Knick and Rotenberry 1998). The ambiguity of absence may cause difficulties for modelers because many modeling methods require absence data. However, many modelers still have used presence-absence model by adding in pseudo-absence locations (Zaniewski et al. 2002; Lutolf et al. 2006; Dillard 2007).

Most current habitat models are enhanced with Geographic Information Science (GIS) products like ArcGIS (ESRI). Values for habitat variables at observed locations are extracted to create the models and then the models are extrapolated to the entire study area (Scott et al. 1993). GIS used in conjunction with habitat models is a popular method for locating and understanding target species (Dye 2003; Gavashelishvili 2004; Sattler et al. 2007; Tarkhnishvili et al. 2009). By extrapolating the results to the whole study area, researchers can analyze and visualize the predicted locations of optimal habitat, locations where the habitat gradually degrades, and locations where the optimum changes to uninhabitable land (Rotenberry et al. 2002; Browning et al. 2005; van Manen et al. 2005). Regardless of method, good habitat models need to be parsimonious and avoid the overfitting that occurs when including every obtainable habitat variable (Boyce et al. 2002). Good models also should be as simple as possible while using the best parameters as predictors—predictors that are relevant to the species with limited reliance on proxies for unmeasured predictors (Anderson et al. 2000; Austin 2002).

1.3.2 Presence-Absence Models

Presence-absence models, usually some version of regression analysis, require locations of where a species exists and locations where the species does not exist. To obtain these data in the field, researchers generally employ one of three sampling designs—random, case-control, and use-availability (Keating and Cherry 2004). The random design randomly selects locations on the landscape that are later surveyed for presence or absence of a species. The case-control design selects presence and absence locations based on previously known areas of use and non-use by a species to ensure a sufficient presence

sample. The use-availability design selects presences based on known areas of use but selects absences randomly on the landscape, leading to the accepted possibility of un-sampled presence locations being classified as absences. Nevertheless, absence data, sampled for or not, carry a high degree of uncertainty, particularly for little-known species. Some species can be elusive, and the lack of observations at a location does not mean the animal never occurs in that location. Also, field work may indicate a species truly is absent from a location, but the absence may not mean the habitat is poor. Perhaps the species has not dispersed to the area yet because of a landscape barrier between current and potential distribution (Boetsch et al. 2003). For the model, absence also should be useful enough to discern between small variations in suitability. A terrestrial plant species, for example, will have no presence locations in water. Allowing a too great a proportion of absences to fall within water may improve the overall goodness-of-fit for the model by predicting between suitable and completely unsuitable habitat. However, the absences would do little to improve the predictive power and detect more subtle nuances across the landscape (Brotons et al. 2004; Engler et al. 2004).

Without true absence locations, modelers have used various techniques to create pseudo-absence locations, such as creating a set of random locations across the landscape (Hirzel et al. 2001; Dillard 2007). When creating random locations modelers must realize similar habitat will likely be in clusters and limit where random locations can occur. Depending on the size of the study area and the number of random locations, a pseudo-absence may overlap fall in perfectly good habitat in the center of a cluster of presence locations. Keating and Cherry (2004) were particularly critical of standard, unmodified models with a use-availability design because of the potential for overlap between

presences and absences. Dillard (2007) attempted to resolve this issue by creating buffers around presence locations of Cheat Mountain salamanders (*Plethodon nettingi*) to prevent random absence locations from falling too close to presence locations. However, Johnson et al. (2006) indicated the amount of contamination, or presence-absence overlap, would have to be rather large to adversely affect the model predictions and slight errors in the model should not preclude its practical use.

1.3.3 Presence-only Models

Presence-only models, such as Mahalanobis Distance (Mahalanobis 1932) and MaxEnt (Phillips et al. 2004), are less common than presence-absence models in the literature but are becoming more popular in species distribution modeling (Browning et al. 2005; Phillips and Dudik 2008; Etherington et al. 2009; Griffin et al. 2010). Presence-only models identify a multi-dimensional, non-uniform space to define habitat from the covariates in the model. The best habitats are locations where the combination of habitat variables is most similar to the observations in value and covariance. Whereas errors in presence-absence models may arise from incorrectly classified absences, errors in presence-only models can be a result of the lack of absence data (Dettmers et al. 2002; Brotons et al. 2004). Rare habitats with presence data may be underestimated because the proportion of presences is much greater in more common habitat. Similarly, other habitats may be overestimated because absence data are not available to offset the proportionality of use (Brotons et al. 2004). These problems primarily arise because researchers often develop presence-only models when absence data are unreliable or because the data were collected at a different time for a different study. While sampling methods designed for the

proposed model may be optimal (Fortin and Drake 2005), researchers have shown that model-specific sampling is not necessary and have produced models that greatly improved efficiency in management (van Manen et al 2005; Etherington et al. 2009).

1.4 Research Objectives

The overall goal of this thesis was to develop habitat models for three salamander species in Great Smoky Mountains National Park. The specific objectives to achieve the overarching goal included investigating the relationship between the salamanders and elements of their habitat, developing a set of habitat models with the use of logistic regression and Mahalanobis distance methods, and comparing the performance of those habitat models. Specifically, I approached my objectives by addressing these main questions:

1. What are the determining factors for salamander habitat in GSMNP at a 30 meter resolution?
2. How do the models support or refute existing knowledge about salamander habitat in GSMNP?
3. For logistic regression models, do case-control models with true absences perform better than use-availability models with pseudo-absences?
4. Can presence-only models built using Mahalanobis distance perform better than presence-absence models using logistic regression models?

1.5 Organization of Thesis

This thesis is divided into five chapters. This first chapter justifies my thesis by explaining the needs of natural resource managers. Additionally, the first chapter provides an overview of relevant literature on salamander species and habitat modeling, concluding with the goals of the study. The second chapter describes the study area and the process of choosing three salamander species out of 30-plus amphibians and the variables to describe them. Chapter Two also describes the two types of mathematical models and the method of comparing models within and between types. Chapter Three presents the results of the study, and Chapter Four discusses those results by examining the model variables and providing comparisons of the models. Chapter Five concludes the thesis and provides suggestions for possible future research.

CHAPTER TWO

2 METHODS

2.1 STUDY AREA

Great Smoky Mountains National Park (GSMNP) was established in 1934 and is the most visited national park in the United States (National Park Service 2010a). The park covers 2,108 km^2 in the Appalachian Mountains, and is found overlapping the southern portion of the border between Tennessee and North Carolina. The elevation ranges from 267 m on the western border of the park at the mouth of Abrams Creek to 2,025 m in the center of the park at the peak of Clingmans Dome. Up to 216 cm of rain per year keep approximately 3,380 km of streams flowing throughout GSMNP. The moist environment is crucial to the 30 salamander species that are known to live within the boundaries of the park (National Park Service 2010b).

The Smoky Mountains are considered one of the most diverse regions in the world for salamander species outside the tropics (Houk 1993; Petranka 1998; Dodd 2004). One of the most popular species in GSMNP is the hellbender (*Cryptobranchus alleganiensis*), a 75-cm long aquatic species that has gills at every life stage. Five families of salamanders occur in GSMNP, but the evolutionary history of the family *Plethodintidae* is the primary reason for the large number of salamander species as they comprise 24 of the 30 species (Highton 1995; Crespi et al 2003; Kozak and Weins 2006).

2.2 Choosing My Target Species

The dataset I used contained locations of individuals of 44 amphibian species collected between 1998 and 2001 by Dr. C. Kenneth Dodd of the University of Florida and a team of researchers. They surveyed at locations throughout the park counting all species of amphibians using various methods such as 10 x 10 m plots and litter bag surveys (Dodd 2004). Each observation contained the species name, coordinates, observer name, site name, and year observed. The data contained no information about the habitat at each site.

Once I obtained the dataset, I made an initial selection of candidate species by excluding species that were not salamanders, such as frogs and newts, and observations without coordinates. Next, I removed species that had fewer than 30 observations to ensure sufficient sample sizes. I then examined research articles, books and field guides to determine the level of existing knowledge for the remaining species. Several species had little to no information available regarding habitat relationships, diet, or species interactions, so I eliminated those species. Because most remaining species belonged to the family Plethodontidae, I removed all other families. For all remaining observations, I used ArcGIS to extract values of habitat variables, including elevation, slope, aspect, disturbance type, understory class, overstory class, and distance to streams.

Ultimately, I chose three plethodontid species—a relative generalist with a broad distribution, a relative specialist with a narrow geographical range, and one species in between the extremes. The terms "generalist" and "specialist" were not intended to categorize species in terms of diet or close associations with other species but refer to extremes of a gradient of geographical range. The three species were *Desmognathus ocoee* (specialist), *Desmognathus conanti* (generalist), and *Plethodon jordani* (in-between). I chose

D. ocoee and *P. jordani* because they are listed as rare species in Tennessee and North Carolina. Given that generalist species are less vulnerable to habitat changes (Sarre et al. 1995; Bentley et al. 2000; Swihart et al 2003), I was not able to use a plethodontid generalist that was also state listed.

2.3 Habitat Variables

For all locations, I characterized habitat using a set of biotic and abiotic variables that were derived from GIS data layers (Table 1). I converted polygon shapefiles to rasters for mapping, spatial calculation, and ease of extrapolating models to the landscape. The converted rasters of categorical variables took the value of the class that covered the largest area in each 30 x 30 m pixel. Each raster also matched the grid layout of the elevation raster to ensure the placement of pixels would not shift in the final output. A few variables, such as Beer's aspect transformation and slope, were derived from elevation using ArcGIS versions 9.3 and 10.

For the categorical variables, I modeled the most prevalent soils, overstories, understories, and disturbance types that were represented in the data (Tables 2–4). The less prevalent types served as reference classes. Specific classes within general variables, such as soil type, were mutually exclusive. Understory was an exception because many types existed sometimes as the sole dominant vegetation and sometimes mixed with other dominant types. Therefore, some locations had more than one understory type at the same time.

Habitat selection for the three salamander species at a scale larger than the home range of an individual but smaller than a regional distribution has not been studied. The

habitat variables I chose to use in the models were a result of what was available and obtainable for GSMNP. Variables were also chosen based on a general knowledge of what may be important to any amphibians, namely moisture, temperature, shelter, and disturbance (Stebbins and Cohen 1995; Petranka 1998). The current lack of habitat information pushed these models to be exploratory analyses to help determine what may define habitat and what variables may be limiting factors at the scale of this study.

Additionally, GSMNP has undergone many changes since the arrival of humans. Fire use by Cherokees earlier than the 1500's encouraged growth of fire-tolerant trees (Delcourt and Delcourt 1997), and fire exclusion, especially over the last century, has created denser understories than would otherwise exist (Harrod et al. 1998). Settlements, such as in Cades Cove and corporate logging in the early 1900's drastically altered much of GSMNP before it was established (Pyle 1988). Fugal infestations, such as chestnut blight, essentially removed all American chestnut trees (*Castanea dentata*) which used to dominate lower elevations (Houk 1993). These recent habitat changes certainly contributed to where some of these salamanders exist today. However, because of the limitation of the dataset used in this analysis, the issues of salamander distribution prior to the establishment of GSMNP as well as potential impacts of climate change were not addressed in this study.

Table 1. Descriptions and value ranges of variables to be used in modeling habitat of the three salamander species.

Variable	Possible Values	Description
Elevation	258–2,024	Elevation [m]
Slope	0–64.1	Slope steepness [°] calculated in ArcGIS
Aspect	0–2	Beers transformation of aspect (Beers et al. 1966)
Distance to stream	0–1,331	Distance to stream [m] calculated from ArcGIS
Understory vegetation	6 categories	Shapefile from CRMS* (2004)
Overstory vegetation	7 categories	Shapefile from CRMS* (2004)
Soil type	5 categories	Shapefile from USDA (2008)
Topographic relative moisture index	0–60	Calculated from 4 slope variables (Parker 1982)
Relative slope position	0–100	Position on a slope a pixel is located (Wilds 1997)
Disturbance History	3 categories	Shapefile based on survey from 1934 (National Park Service 2007)

***** Center for Remote Sensing and Mapping Science

Table 2. Overstory vegetation classes and area statistics for GSMNP.

Overstory Vegetation	Area (km²)	Percent
Submesic to Mesic Oak-Hickory Forest	446.3	21.2
Southern Appalachian Northern Hardwood Forest	445.6	20.1
Southern Appalachian Cove Hardwood Forest	353.4	16.8
Subxeric to Xeric Chestnut Oak/Hardwood Forest/Woodland	332.5	15.8
Xeric Pine Woodland	185.0	8.8
Red Spruce-Fraser Fir Forest	56.7	2.7
Montane Alluvial Forest	26.8	1.3

Table 3. Generalized understory vegetation and area statistics for GSMNP.

Understory Vegetation	Area (km²)	Percent
Herbaceous and deciduous understory	1,028.8	48.8
Rhododendron understory	765.8	36.3
Kalmia understory	350.5	16.6
Hemlock understory	320.3	15.2
Pine understory	265.9	12.6
Spruce understory	94.7	4.5

Table 4. Generalized soil types and area statistics for GSMNP.

General Soil Units	Area (km²)	Percent
Frigid Anakeesta Slate: Luftee soils	70.2	3.3
Frigid Sandstone: Breakneck or Oconaluftee soils	296.5	14.1
Mesic Metasandstone: Ditney or Soco soils	992.3	47.1
Mesic Silstone: Junaluska	212.6	10.1
Large Basins of Colluvium: Spivey soils	235.3	11.2

2.4 Habitat Modeling Methods

I constructed three models for each species—two presence-absence models and one presence-only model—for a total of nine models. All models were developed based on data with a resolution of 30 x 30 m. The first presence-absence model for each species was a use-availability model based on the sampling design from Keating and Cherry (2004). Each model had a random set of 142 pseudo-absence locations to capture habitat availability throughout GSMNP. The same 142 pseudo-absences were used for all three species. I used the ArcMap tool, Create Random Points, to obtain this random set of absences. I call these models "use-availability logistic regression". The second presence-absence modeling approach followed the case-control sampling design from Keating and Cherry (2004). By using survey sites where the target species was not observed, I selected a set of true absences for each species approximately equal to the number of presences for each species. I call these models "case-control logistic regression". The presence-only models required no absence data and are called by their method "Mahalanobis distance". All models assume that the species are distributed optimally in their environment, the data collection has captured that distribution, and the observations are independent of each other.

I used logistic regression in a stepwise fashion for both presence-absence models. Logistic regression is useful when the dependent variable is binary (Hosmer and Lemeshow 1989). Specifically, for this study, the dependent variable was either presence (1) or absence (0). Different from linear regression, no assumption is made about the distributions of the variables in logistic regression (Bewick et al. 2005). The basis of logistic regression is the logit function which produces the log of the odds ratio of a particular outcome, in this case good salamander habitat:

$$\ln\left(\frac{P}{1-P}\right) = \sum a_n x_n + b$$

where *x* is the value of habitat variable *n*, *a* is the regression coefficient for habitat variable *n*, and *b* is log odds estimate when all variables equal zero (Aldrich and Nelson 1982; Peng et al. 2002). The equation can be used to obtain a probability value by taking the antilog of both sides and solving for *P*:

$$P = \frac{1}{1 + e^{-(\sum a_n x_n + b)}}$$

The probabilities given henceforth represent the likelihood that a pixel was habitat for a salamander species given a set of habitat variables. The model parameters a_n and b in the logistic regresson model were determined by a maximum likelihood estimation (MLE). Much like how ordinary least squares optimizes the values of model parameters to minimize the sum of the squared differences between the observed and predicted *y* values, MLE optimizes the values of the parameters by maximizing the probability value for each observation (Hosmer and Lemeshow 1989). The models with optimized parameters could then be used to determine the probability of habitat use for each species when extrapolated to GSMNP.

For the presence-only models, I calculated Mahalanobis distance (D^2). Mahalanobis distance measurements result in values of dissimilarity where smaller values are more similar to the sample locations with respect to the mean and covariance of the measured habitat variables. Based on a covariance matrix, the D^2 values are normalized and are thus not affected by the scale of the input variables (Clark et al. 1993). Mahalanobis distance is calculated using the following formula:

$$D^2 = (x - \mu)' \sum^{-1} (x - \mu)$$

where x is a vector of the variables for any given location, μ is a mean vector of the variables of all observed locations, and Σ^{-1} is the inverse of the variance-covariance matrix of the sample locations (Rao 1952). In the variance-covariance matrix, the diagonal values are the variances of the variables and give weights to the individual variables, whereas the off-diagonal values give weights to the various relationships between variables (Clark et al. 1993; Feng et al. 2009). Also, by retaining the squared difference between a location and the mean vectors of the observed locations, the models place greater weight on points farther away from the mean.

I used both continuous and categorical variables to calculate Mahalanobis distance values for each observation and each pixel in GSMNP. Once I calculated D^2 for all locations in GSMNP, I generated random locations approximately equal to the number of observations to evaluate discrimination in the models. I classified salamander habitat by choosing distances that had a lower dissimilarity value than locations randomly available on the landscape. The greatest difference between the cumulative frequency distributions of each species and the random locations, based on their D^2 values, was chosen as the threshold that identified habitat for the species (Pereira and Itami 1991; Feng et al. 2009).

2.5 Choosing the Best Models

The first step in choosing the best logistic regression model for each species was to determine if each variable was significant ($p<0.05$) in a univariate model. If a variable was not significant on its own, I did not include it in the final model. Secondly, I ranked all models using Akaike's Information Criterion (AIC) to select the best univariate model for

each species (Burnham and Anderson 1998; Anderson et al. 2000). AIC ranks models based on the disagreement between the models and the real world which is estimated using the log-likelihood and the number of parameters in each model (Akaike 1973; Stanley and Burnham 1998). AIC is a commonly used technique to select the best logistic regression model among a group of possibilities (Glenn et al. 2004; Karpanty et al. 2006; Dillard et al. 2008). For consistency, I ranked each logistic regression model by AIC corrected for small samples (AIC_c).

My third step was to examine combinations of variables. Burnham and Anderson (1998) suggested an *a priori* approach to modeling instead of the data dredging that usually follows exploratory analyses. While I was limited by available literature, the variables I chose were not without biological basis. Additionally, Burnham and Anderson (1998) conceded that ranking models based on an information-theoretic criterion, such as AIC, may be a better approach in exploratory analysis than using an approach based solely on statistical significance. The variable with the top ranked AIC_c of the univariate models was included first and the remaining variables were added iteratively. I only added multivariate models to the AIC_c ranking table in which the confidence intervals of the exponent of the parameter estimates did not cross one. The best model was the one with the lowest AIC_c score unless a more parsimonious model had an AIC_c score within 2.0 of the best model (Burnham and Anderson 1998).

Once I chose the best models for both logistic regression methods, I compared the models by evaluating their discrimination and calibration. Discrimination is the ability of a model to correctly identify between presence and absence observations (Pearce and Ferrier 2000). A simple comparison is to use a classification table to determine the

proportion of correctly predicted observations. A second test of discrimination is to calculate the area under the Relative Operating Characteristic (ROC) curve. The ROC curve is a graph of true positive versus false positive predictions and is an effective way to compare different models (Pearce and Ferrier 2000). The larger the area under the ROC curve (AUC), the better the model, with an AUC value of 1.0 representing a perfect model.

Calibration differs from discrimination in that it measures the ability of a model to predict probability in agreement with the observed proportion of presences (Pearce and Ferrier 2000). For example, a test of calibration expects that out of locations with predicted probability of approximately 0.10, 1 in 10 will be presence locations. Apart from a perfect model, there will be disagreement between the predicted probabilities and the observed proportions. Disagreement can be split into three groups: bias, spread, and unexplained error. Bias is a consistent under- or overestimation of the model. On a calibration diagram of predicted versus observed values, bias moves the regression line up or down parallel to the expected $y=x$ line. Spread is an alteration of the regression line from the expected $y=x$. A model with spread error has a regression line with a slope different from 1, meaning it both under- and overestimates at different places in the model. Unexplained error is imprecise predictions from variation not accounted for with variables included in the model. I was able to visualize bias and spread by graphing the observed proportions versus predicted probability. I split probability into five classes, plotted the medians of each class, and used the resulting linear regression as a measurement of bias and spread (Pearce and Ferrier 2000).

With Mahalanobis distance, I was limited in how much I could depend on statistical significance or information-theoretic approaches to determine if a variable performed well

(Johnson and Gillingham 2005). Mahalanobis distance deals with variables not only as independent predictors but also as interacting pairs, making a completely objective measure of variable quality difficult to achieve. Furthermore, variable selection for D^2 models mathematically is largely absent from the literature, leading me to a more subjective approach. I first removed any categorical variables that comprised less than 10% of all observations as those variables already show a lack of importance compared to others. Next, I calculated the Mahalanobis distance values for each observation based on the remaining variables and determined the habitat threshold value by comparing the model with random cumulative frequencies. To evaluate the prediction ability of the D^2 models, I used a ten-fold cross-validation which split the data into ten randomly generated groups of equal size (Fielding and Bell 1997). By removing one group at a time, I created ten D^2 validation models. Each validation model was tested by the one-tenth of data each model omitted. The percentage of tested observations that remained under the original habitat threshold value gave the original model a measure of accuracy.

Poor cross-validation results of the original D^2 models for each species led me to improve the models with a further reduction in the number of variables. Specifically, I made model improvements by analyzing the means of the continuous variables and the proportions of the categorical variables. I performed a statistical t-test on the distribution of each continuous variable by comparing observed locations to a random selection of available locations. If the mean of a continuous variable was not significantly different than the mean of the random locations, then that variable was a candidate for removal. If a categorical variable was represented in a smaller proportion in favorable habitat than was available in GSMNP, then it was a candidate for removal. Based on the assumptions that the

salamanders are optimally distributed in GSMNP and that the variables in a presence-only model contain observed locations, a habitat variable containing a salamander observation cannot be ecologically detrimental to the species. Therefore, removing categorical variables from the model that were in predicted favorable habitat at a lower proportion than in GSMNP removed the possibility variables were being harshly punished and effectively classified as an absence even though they contained a presence location. Those variables were instead relegated to the reference classes of variables within the same category or excluded from the model all together. For example, if a model contained soil type A and B, and soil B was removed, it would remain in soil A's reference class but treated the same as any soil not A in the model. If a model contained only soil type B and it needed to be removed, soil type would not be considered in the model at all, being deemed unimportant. After reducing the number of variables, I created new models for each species, determined a new threshold value, cross-validated the results, and calculated the D^2 for every pixel in GSMNP.

For the best model choice between methods, I compared the classification rates by using the presence-absence data from the case-control model. Because the case-control model used observed presence and observed absence locations it was a good measure of predictive success between the models. Depending on the land manger using the results, however, different estimates of habitat may be desired. Therefore, the classification results were combined with subjective evaluations to determine the benefits of each model for each species.

CHAPTER THREE

3 RESULTS

3.1 *DESMOGNATHUS CONANTI*

3.1.1 USE-AVAILABILITY LOGISTIC REGRESSION

Of the logistic regression models created, the best model included elevation, historically settled areas, pine understory, and flood overstory (Table 5). The weight of evidence (w_i = 0.78) was nearly four and a half times greater for the best model than the second best model. Presence of the spotted dusky salamander was positively associated with historically settled areas, pine understory, and floodplain overstory. The salamander was negatively associated with increasing elevation and predicted habitat generally did not occur above 1000 m. The next best model included elevation, historically settled areas, and flood overstory and received a moderate amount of support (ΔAIC_c = 3.0; w_i = 0.18). This model indicated the salamander was positively associated with historically settled areas and floodplain overstory while being negatively associated with increasing elevation. The remaining multivariate and univariate models received far less support than the top two models (ΔAIC_c > 7.9).

The best model had a classification accuracy of 71.8% and an area under the ROC (AUC) of 0.83 (Figure 1). That model had a small amount of spread error (slope = 0.98), underestimating slightly from 0.2 to 0.4 and overestimating from 0.4 to 0.7 (Figure 2). The model had a small amount of positive bias (y-int = 0.01). The second-best model had an accuracy of 71.4% and an AUC of 0.81. However, the model had a relatively large amount of spread error (slope = 0.39) and a large bias (y-int = 0.29). The equation for logistic

regression from the best model was used to map the model across the landscape of GSMNP (Figure 3), and about 655 km^2 of GSMNP was predicted habitat for the spotted dusky salamander.

Table 5. Selection for the use-availability logistic regression models for *Desmognathus conanti*.

Parameters	K	AIC_c	% Cor	R^2	Delta	w
elevation, settled, pine understory, floodplain overstory	4	276.6	71.8	0.42	0	0.78
elevation, settled, floodplain overstory	3	279.6	71.4	0.40	3.0	0.18
elevation, settled	2	284.5	71.8	0.37	7.9	0.01
elevation, floodplain overstory	2	287.6	71.4	0.36	11.0	0.00
elevation, distance to stream	2	289.6	72.6	0.35	13.0	0.00
elevation, undisturbed	2	292.4	72.6	0.34	15.7	0.00
elevation	1	292.7	71.4	0.34	16.0	0.00
settled	1	338.5	66.5	0.15	61.9	0.00
undisturbed	1	339.6	57.5	0.14	63.0	0.00
distance to stream	1	343.1	58.6	0.13	66.5	0.00
northern hardwood overstory	1	351.0	54.9	0.09	74.3	0.00
pine understory	1	353.2	61.3	0.08	76.6	0.00
slope	1	353.7	59.8	0.08	77.1	0.00
floodplain overstory	1	355.0	57.9	0.07	78.4	0.00
Junaluska soil	1	356.6	60.5	0.06	79.9	0.00
heavily disturbed	1	357.7	55.3	0.06	81.1	0.00
Ditney/Soco soil	1	362.4	57.1	0.04	85.7	0.00
Kalmia understory	1	364.8	57.1	0.02	88.1	0.00
herbaceous understory	1	365.4	55.6	0.02	88.8	0.00

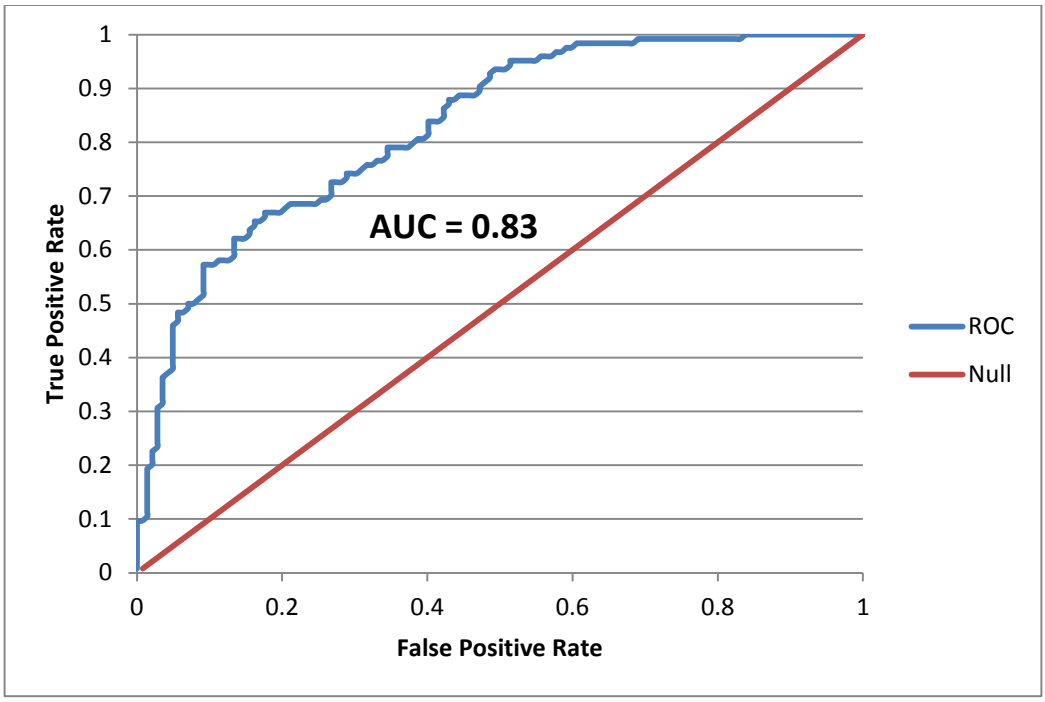

Figure 1. Relative Operating Curve for the best use-availability logistic regression model for *D. conanti*.

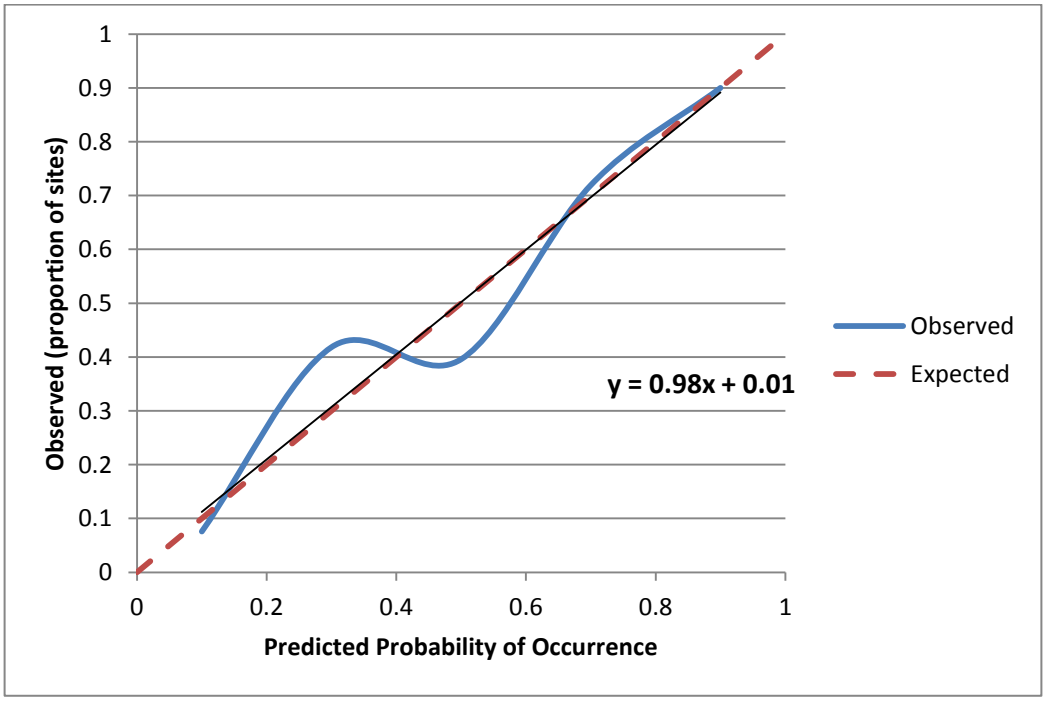

Figure 2. Calibration curve (observed) for the best use-availability model for D. conanti including the expected values and the trendline for the calibration curve.

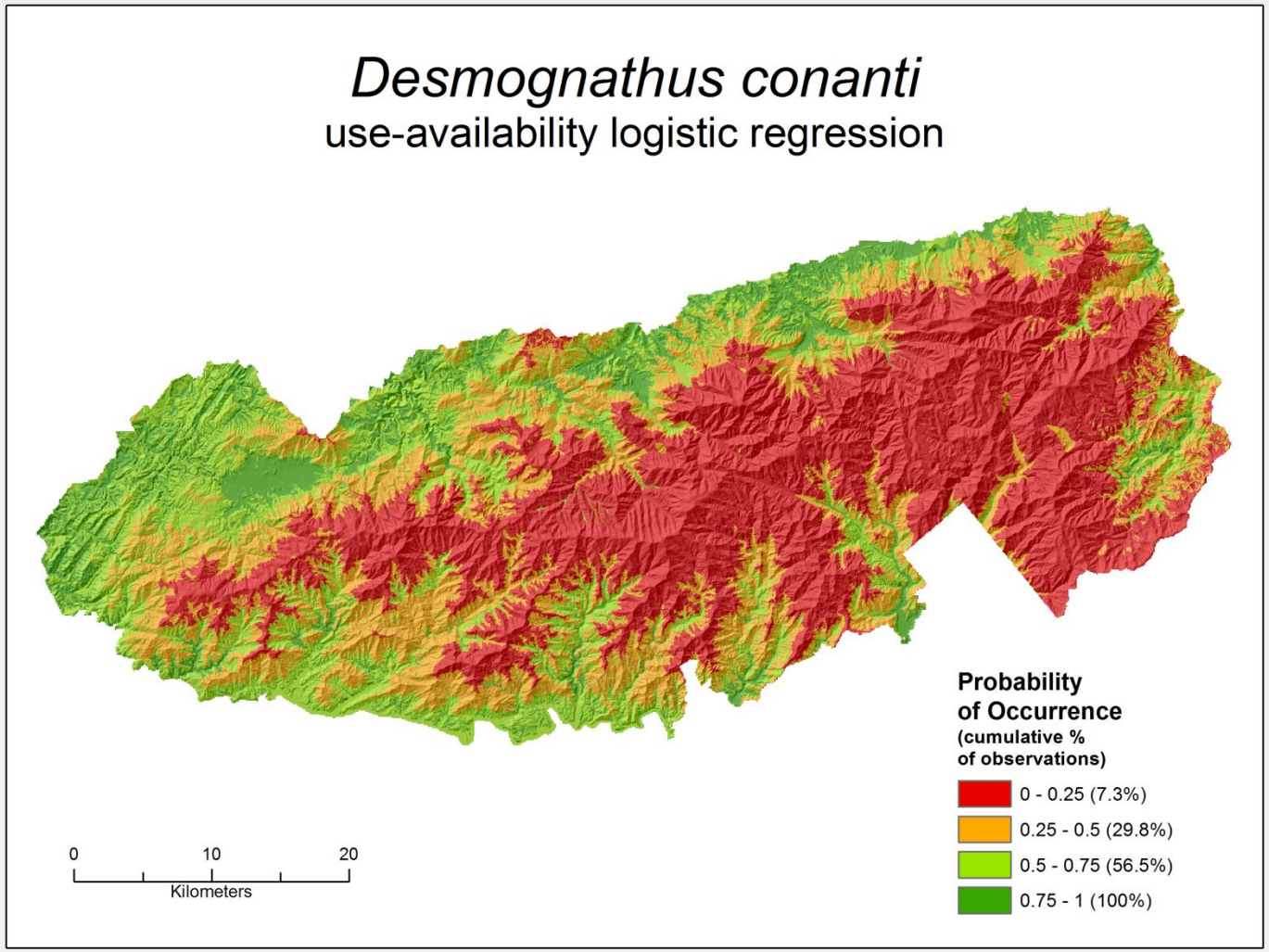

Figure 3. Use-availability logistic regression model extrapolation for *Desmognathus conanti.*

3.1.2 Case-Control Logistic Regression

The absence locations in this model were generated using other survey sites in the study area. The best model included elevation and historically undisturbed areas and had a weight of evidence approximately 17 times greater than the second-best model (w_i = 0.81; Table 6). Presence of the spotted dusky salamander was negatively associated with both elevation and historically undisturbed areas. The next best model included elevation and northern hardwood overstory but did not receive much support (ΔAIC_c = 5.6; w_i = 0.05). This model indicated the salamander was negatively associated with elevation and northern hardwood overstory. Two other models also received limited support, including the model with elevation as the lone variable (ΔAIC_c = 5.7; w_i = 0.05). All the univariate models received far less support (ΔAIC_c > 25.7).

The best AIC model had a classification accuracy of 70.0% and an AUC of 0.73 (Figure 4). The best model had a small amount of spread error (slope = 1.10), overestimating probabilities < 0.3 and around 0.7 and underestimating probabilities between 0.3 and 0.7 (Figure 5). The model had a smaller amount of negative bias (y-int = –0.031), pulling all the probabilities down slightly. The second-best model had an accuracy of 68.3% and an AUC of 0.72. It had spread error (slope = 1.04) and bias (y-int = –0.01) nearly identical to the best model with a discernible difference in the smaller degree of underestimation from 0.3 to 0.7. The logistic regression equation from the best model was used to calculate the probabilities of occurrence over all of GSMNP (Figure 6). Overall, 1,076 km^2 of GSMNP was predicted to be spotted dusky salamander habitat.

Table 6. Selection for the case-control logistic regression models for *Desmognathus conanti*.

Parameters	K	AIC_c	% Cor	R^2	Delta	w
elevation, undisturbed	2	280.2	70.0	0.29	0	0.81
elevation, northern hardwood overstory	2	285.8	68.3	0.27	5.6	0.05
elevation	1	285.9	68.3	0.26	5.7	0.05
elevation, northern hardwood overstory, distance to stream	3	285.8	69.5	0.28	5.7	0.05
elevation, distance to stream	2	286.0	68.7	0.27	5.8	0.00
undisturbed	1	305.9	62.1	0.17	25.7	0.00
northern hardwood overstory	1	306.8	62.6	0.16	26.6	0.00
distance to stream	1	317.5	60.1	0.11	37.2	0.00
settled	1	328.7	59.3	0.05	48.5	0.00
Junaluska soil	1	328.7	57.2	0.05	48.5	0.00
floodplain overstory	1	332.0	53.1	0.04	51.8	0.00
Dellwood soil	1	333.0	52.7	0.03	52.8	0.00
pine understory	1	333.7	55.1	0.03	53.5	0.00
slope	1	334.3	54.7	0.02	54.1	0.00
herbaceous understory	1	334.4	56.8	0.02	54.2	0.00
heavy disturbance	1	334.4	55.1	0.02	54.2	0.00

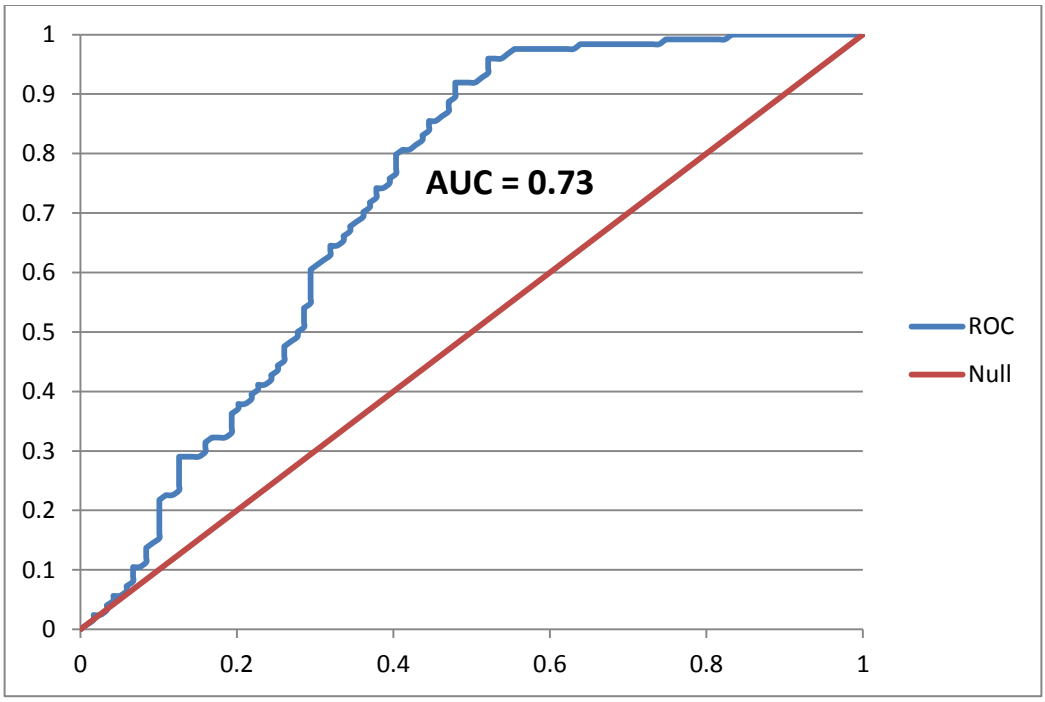

Figure 4. Relative Operating Curve for the best case-control logistic regression model for *D. conanti*.

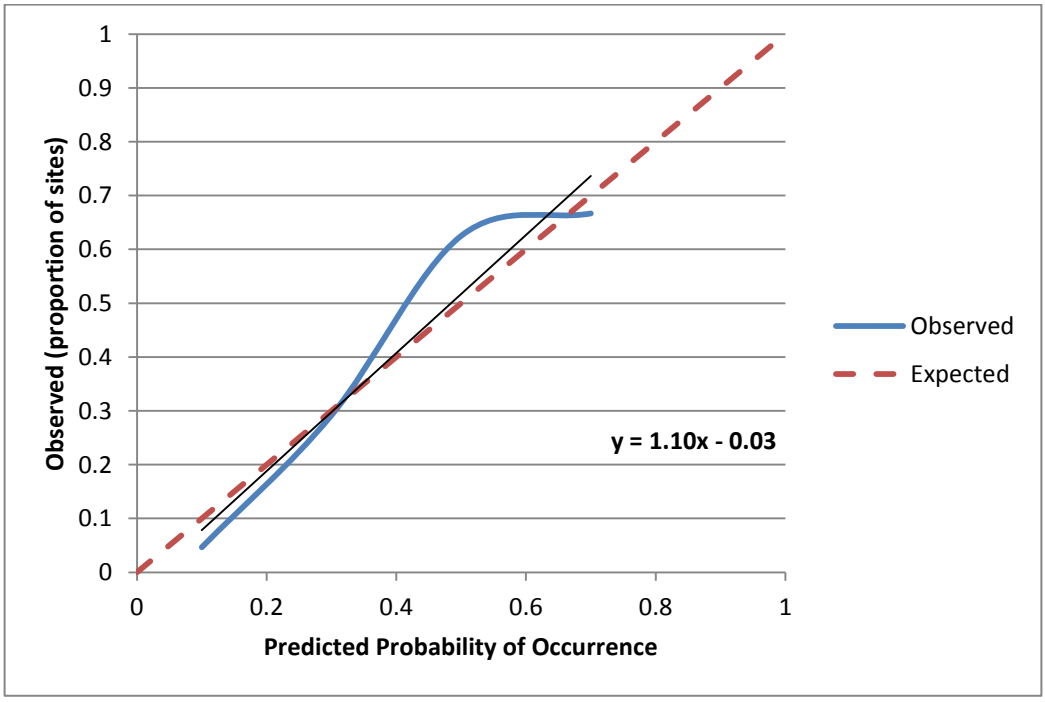

Figure 5. Calibration curve (observed) for the best case-control model for *D. conanti* including the expected values and the trendline for the calibration curve.

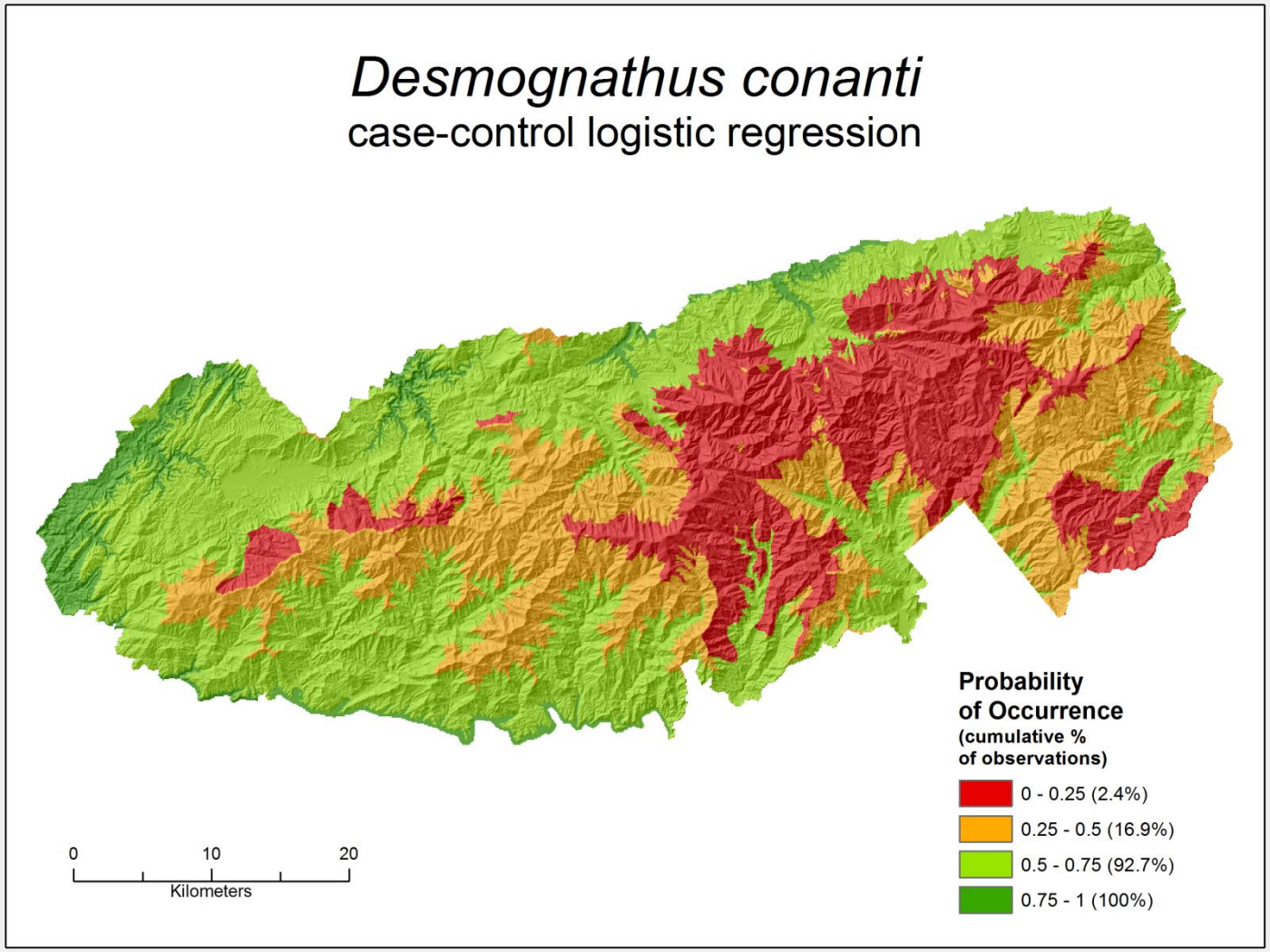

Figure 6. Case-control logistic regression model extrapolation for *Desmognathus conanti.*

3.1.3 MAHALANOBIS DISTANCE

The data for the spotted dusky salamander included 124 unique locations inside GSMNP. The mean values for each parameter are listed in Table 7. The original model had 16 parameters with Mahalanobis distance values for GSMNP that ranged from 6.2 to 230.5 ($\bar{x}$ = 24.5, SD = 14.3). The 124 spotted dusky salamander locations had D^2 values that ranged from 7.6 to 42.6 with a mean of 16.1 (SD = 4.3). Conversely, the 124 random (null model) locations had D^2 values that ranged from 8.6 to 93.0 ($\bar{x}$ = 25.3, SD = 12.9). The difference between the distributions of the observed and random locations was significant ($p < 0.01$).

Using the cumulative frequency distributions of the D^2 values of the observed locations and random locations, the greatest distance between the two cumulative frequency graphs occurred at a D^2 value of 19.4, which I initially chose as the threshold between favorable and unfavorable habitat (Figure 7). However, ten-fold cross-validation indicated the D^2 model could be improved. In the cross-validation, the percentage of observations that stayed under the favorable habitat threshold of 19.4 was only 57.3%. I subsequently removed six parameters to improve model performance. Similar mean values for continuous variables and smaller proportions in favorable versus larger in unfavorable habitat were determining factors for removal (Table 8).

Table 7. Mean values for variables of *Desmognathus conanti* locations (* indicates variable was used in final model).

Variable	Mean	Variable	Mean
Elevation*	672.5	hemlock understory	0.2
Slope*	17.6	*Kalmia* understory	0.2
RSP	49.8	cove overstory*	0.2
TRMI	29.1	chestnut-oak overstory	0.1
historically settled*	0.4	flood overstory	0.1
Ditney-Soco soils*	0.3	oak-hickory overstory*	0.3
Junaluska soils*	0.3	pine overstory*	0.1
Spivey soils*	0.2	distance to stream*	105.5

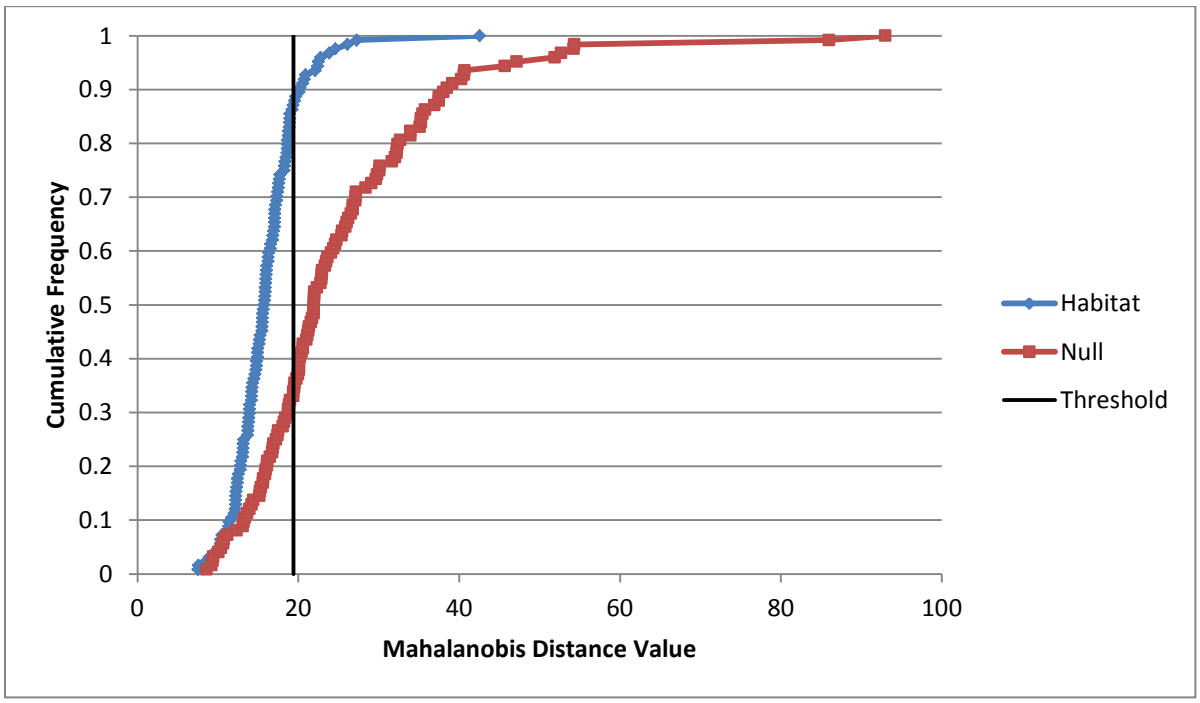

Figure 7. Cumulative frequency graph for the original Mahalanobis distance model versus the random (null) model for *Desmognathus conanti.*

Table 8. Favorable versus unfavorable habitat based on the original model for *D. conanti*. Continuous variables are compared by mean and categorical variables are compared by their proportion (* indicates variable used in final model).

	Favorable		Unfavorable	
Variable	**Mean**	**SD**	**Mean**	**SD**
Elevation*	830.2	237.1	1,156.9	344.7
Slope*	21.3	8.9	24.0	9.2
RSP	48.7	41.5	51.5	42.1
TRMI	28.3	11.2	27.3	11.0
Distance to stream*	125.7	86.7	241.9	170.9
	Prop		**Prop**	
Historically settled*	0.21		0.13	
Ditney-Soco soils*	0.58		0.41	
Junaluska soils*	0.16		0.06	
Spivey soils*	0.15		0.08	
Hemlock understory	0.11		0.18	
Kalmia understory	0.12		0.18	
Cove overstory*	0.25		0.10	
Chestnut-oak overstory	0.12		0.17	
Floodplain overstory	0.01		0.01	
Oak-hickory overstory*	0.32		0.12	
Pine overstory*	0.09		0.07	

The improved, final model contained 10 of the original 16 parameters and had distances ranging between 4.8 and 201.7 ($\bar{x}$ = 17.8, SD = 13.8). The observed salamander locations had distances between 5.2 and 30.7 ($\bar{x}$ = 9.9, SD = 3.8) compared to null model distances between 5.3 and 82.0 ($\bar{x}$ = 16.8, SD = 12.5). The means of the observed and random locations were significantly different ($p < 0.01$). Again, I used the cumulative frequency distributions of the D^2 values of the observed locations and random locations to determine value of the threshold of favorable habitat. The threshold to determine *D. conanti* habitat was at a D^2 value of 13.5 (Figure 8). Of all 124 spotted dusky salamander locations, 91.9% of the locations had D^2 values below 13.5. Other thresholds used to more easily visualize spotted dusky salamander habitat were $D^2 < 7.2$ (25% of salamander observations), $D^2 < 9.3$ (50%), $D^2 < 13.5$ (91.9%), $D^2 < 16.2$ (95%), $D^2 < 30.7$ (100%). Pixels in GSMNP with values greater than 30.7, where no salamanders were observed, constituted the sixth class.

By extrapolating the D^2 model onto the GSMNP landscape, I estimated 1,071 km^2 of spotted dusky salamander habitat out of 2,108 km^2 of area available in GSMNP (Figure 3). Favorable habitat was mostly continuous on the western side of the park, with limited habitat in the higher elevations on the eastern portion of the park. Elevation and nearness to streams were the two most important variables based on four measurements: (1) the means of the locations were different from means of GSMNP ($p < 0.01$); (2) coefficients of variation for both variables were small in observed locations compared with both variables across GSMNP; (3) coefficients of variation were small compared with the coefficients of variation for other variables in the model (Table 9); (4) both variables showed strong correlation with numerous other variables in the model. Elevation in favorable habitat was

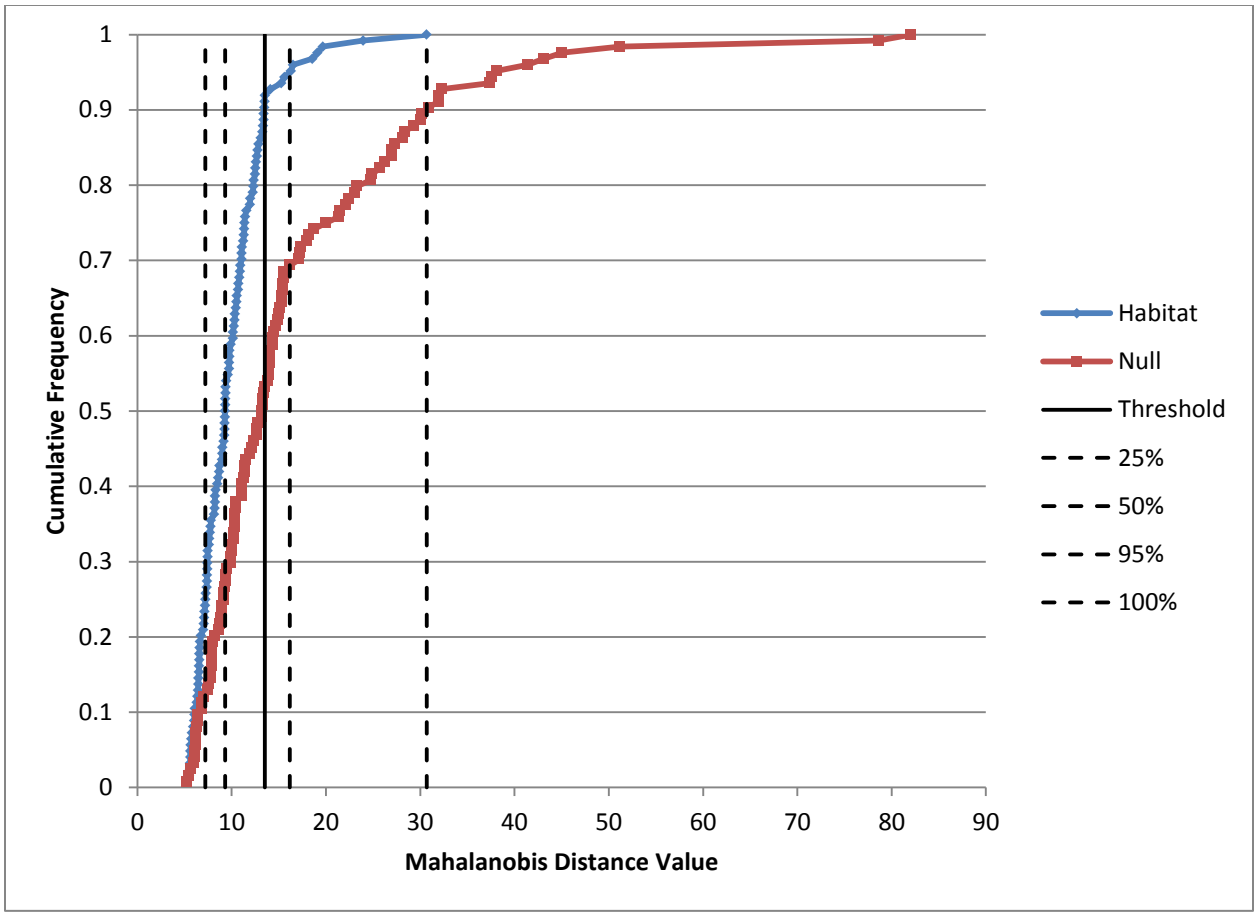

Figure 8. Cumulative frequency graph for the improved Mahalanobis distance model versus the random (null) model for *Desmognathus conanti.*The solid line is the threshold between good and bad habitat. The dashed lines represent the cumulative percentage of salamander observations below each line.

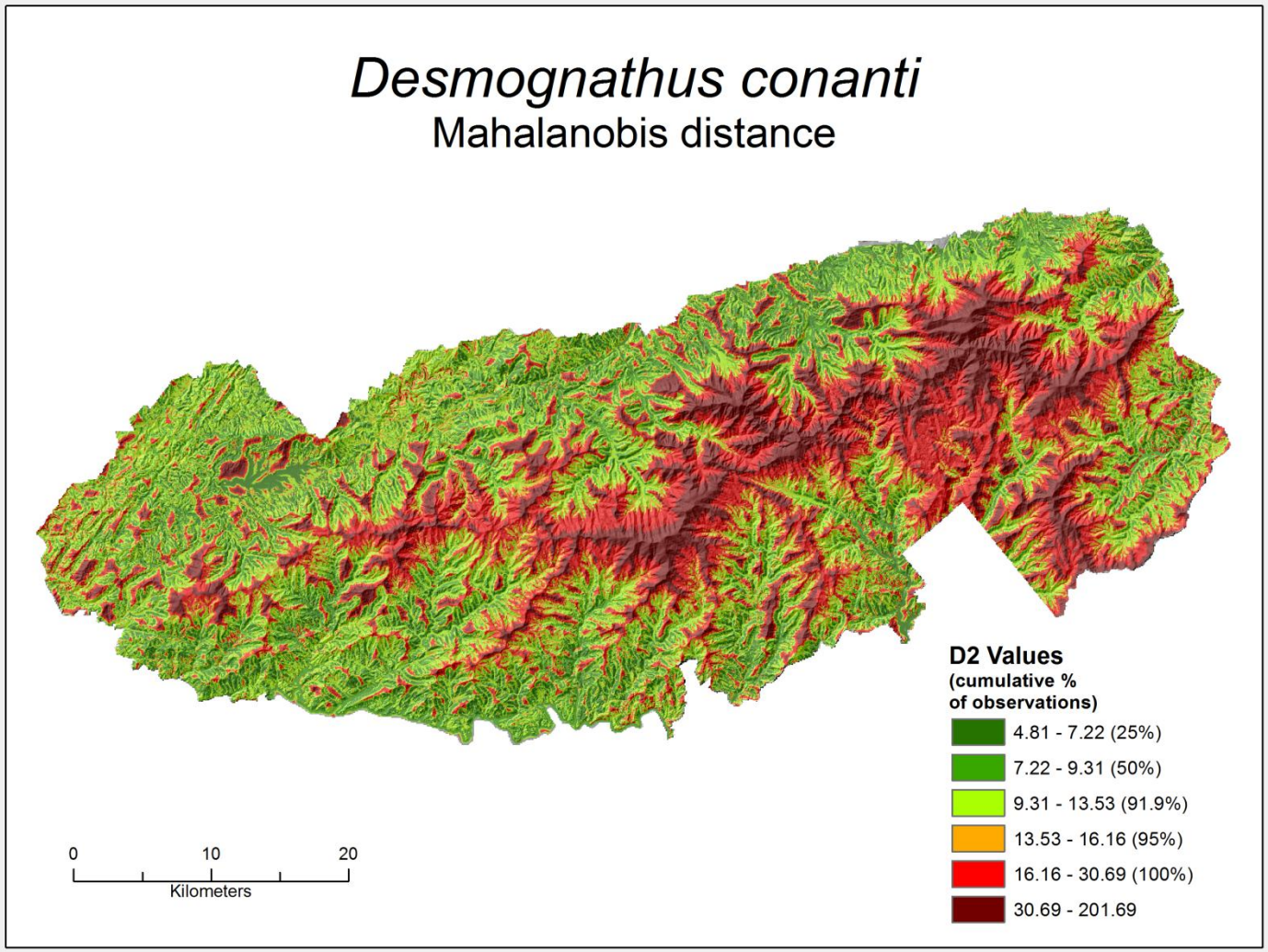

Figure 9. Improved Mahalanobis distance model for *Desmognathus conanti.*

Table 9. Coefficients of variation for favorable habitat compared to GSMNP as a whole for each variable used in the *D. conanti* Mahalanobis distance model.

Variable	**Favorable**	**GSMNP**
Elevation	0.32	0.35
Slope	0.61	0.41
Historically settled	1.14	2.31
Ditney/Soco soil	1.54	1.04
Junaluska soil	1.67	3.00
Spivey soil	2.05	2.91
Cove overstory	2.00	2.31
Oak-hickory overstory	1.74	2.00
Pine overstory	2.61	3.50
Distance to streams	0.77	0.80

lower ($\bar{x}$ = 814.3, SD = 229.0) than in unfavorable habitat ($\bar{x}$ = 1230.5, SD = 315.0). Also, favorable habitat was closer to streams ($\bar{x}$ = 123.8, SD = 83.7) than unfavorable habitat ($\bar{x}$ = 265.0, SD = 173.5). In fact, areas of unfavorable habitat at lower elevations were relatively far away from streams (about 150–200 m away). Elevation seemed to be the limiting factor at about 1,100 m with only five observed locations above that elevation. Additionally, at higher elevations the salamanders may be more terrestrial, avoiding close proximity with streams ($\rho = 0.16$, $p = 0.07$).

Other variables in the model were slope, historically settled areas, soils (Ditney/Soco, Junaluska, and Spivey), and overstories (cove, oak-hickory, and pine). Ditney/Soco soils are metasandstone and both are largely rock outcrops in oak-hickory and yellow pine forests (USDA 2009). However, those soils and overstory types were not correlated in the observed data. Nearly half the park (48.2%) is a Ditney or Soco soil, but 59.5% of the predicted habitat was a Ditney/Soco soil compared with 35.8% of unfavorable habitat, suggesting possible selection for this soil type. Junaluska soils are mostly siltstone but have colluviums unrelated to the underlying siltstone that consist of metasandstone from the Ditney/Soco soils. Oak-hickory and yellow pine also dominate the forests on Junaluska soils with tulip poplar, *Liriodendron tulipifera*, most prominent on the colluvial areas. Approximately 10.2% of GSMNP is a Junaluska soil, but the soil comprised 15.3% of predicted habitat versus 4.7% of unfavorable habitat. The Ditney/Soco soils in colluvium may be driving the relationship with Junaluska soils. Spivey soils are mesic with the possibility of many seeps. *L. tulipifera* also dominates this soil type. 15.1% of favorable habitat was a Spivey soil while only 7.5% of unfavorable areas were Spivey soils. Overall, *D. conanti* was more likely to be in Ditney/Soco soils at higher elevations ($r_{pb} = 0.25$, $p < 0.01$)

and in Junaluska soils at lower elevations (r_{pb} = –0.26, p < 0.01). The salamander was also more likely to be in Spivey soils in areas closer to streams (r_{pb} = –0.28, p < 0.01) whereas locations farther from streams were more likely to be in either Ditney/Soco (r_{pb} = 0.20, p < 0.03) or Junaluska soils (r_{pb} = 0.20, p = 0.02). Similarly, the salamander was more likely to be in cove forests—dominated by trees such as *L. tulipifera*, *Fagus grandifolia*, and *Acer sp.*—as elevation increased (r_{pb} = 0.18, p < 0.05). Cove forest was present in 20.4% of favorable habitat versus 12.3% of unfavorable areas.

Lastly, slope was lower in favorable habitat ($\bar{x}$ = 21.1, SD = 8.8) versus unfavorable areas ($\bar{x}$ = 24.7, SD = 9.7) and had relatively strong correlations with Ditney/Soco soils (r_{pb} = 0.34, p < 0.01) and increased distance from streams (r = 0.16, p = 0.07). Settled areas comprised 20.1% of favorable habitat versus only 11.9% of unfavorable areas. Locations in historically settled areas were likely to be closer to streams (r_{pb} = –0.19, p < 0.04) and have lower elevations (r_{pb} = –0.16, p < 0.07). Oak-hickory forests were represented in 20.4% of favorable habitat against 7.5% in unfavorable areas but were only correlated with the other overstory variables. Pine overstory was likewise uncorrelated with most other variables except for historically settled areas (ϕ = –0.24, p < 0.01).

A ten-fold cross-validation showed much improvement over the original model. The percentage of observations that stayed under the favorable habitat threshold of 13.5 ranged between 61.5% and 100% across the ten groups. Overall, the average was 77.4% of observations that stayed within the threshold of favorable habitat versus 57.3% in the original model. I am more confident in the results of the improved, final model based on this outcome.

3.2 *DESMOGNATHUS OCOEE*

3.2.1 USE-AVABAILBILITY LOGISTIC REGRESSION

The best use-availability logistic regression model included elevation and spruce understory (Table 10). Presence of the Ocoee salamander was positively associated with higher elevation but was negatively associated with spruce understory. The model did not predict locations below approximately 1,500 m but also punished areas that were both high elevation and spruce understory. The weight of evidence (w_i = 0.97) for the best model was over 27 times greater than the second best model. The next best model included elevation only and received a little support (ΔAIC_c = 6.6; w_i = 0.04). This model included only a single parameter and indicated the salamander was positively associated with increasing elevation. The remaining multivariate and univariate models received far less support than the top two models (ΔAIC_c > 19.2).

The model with elevation and spruce understory classified 89.0% of presences and absences correctly and had an AUC of 0.90 (Figure 10). Approximately 121 km^2 of GSMNP was considered favorable habitat for the Ocoee salamander. The spread error for the best model was small (slope = 1.15), overestimating between 0 and 0.5 and underestimating a similar amount from 0.5 to 1.0 (Figure 11). The model also had a small amount of negative bias (y-int = –0.09), pulling probabilities down slightly. The single parameter, "elevation", model had an accuracy of 87.4% and an AUC of 0.885. The spread error (slope = 1.196) caused the model to overestimate more often and the larger negative bias (y-int = –0.101) caused the probabilities to be lower overall than the best model. Therefore, the equation for the "elevation and spruce understory" model was used to calculate the probability of occurrence for the Ocoee salamander across the landscape of GSMNP (Figure 12).

Table 10. Selection for the use-availability logistic regression models for *Desmognathus ocoee*.

Parameters	K	AIC_c	% Cor	R^2	Delta	w
elevation, spruce understory	2	118.6	89.0	0.53	0	0.96
elevation	1	125.2	87.4	0.48	6.6	0.04
Breakneck/Oconaluftee soils	1	162.0	80.8	0.25	43.4	0.00
northern hardwood overstory	1	173.0	78.0	0.17	54.4	0.00
Ditney/Soco soils	1	177.1	78.0	0.13	58.5	0.00
undisturbed	1	178.2	78.0	0.13	59.6	0.00
distance to stream	1	181.8	80.2	0.10	63.1	0.00
cove forest overstory	1	187.2	78.0	0.05	68.6	0.00
light disturbance	1	187.2	78.0	0.05	68.6	0.00
spruce-fir overstory	1	187.9	78.0	0.05	69.2	0.00
oak-hickory overstory	1	188.5	78.0	0.04	69.9	0.00
selective disturbance	1	188.8	78.0	0.04	70.1	0.00
spruce understory	1	190.5	78.0	0.21	71.9	0.00

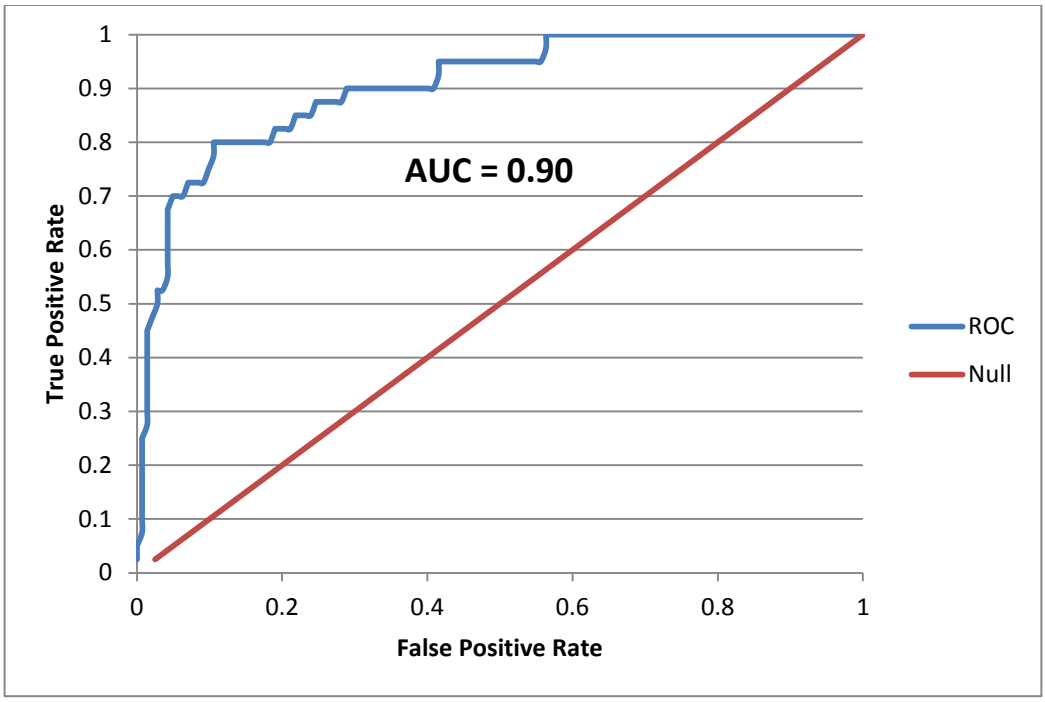

Figure 10. Relative Operating Curve for the best use-availability logistic regression model for *D. ocoee.*

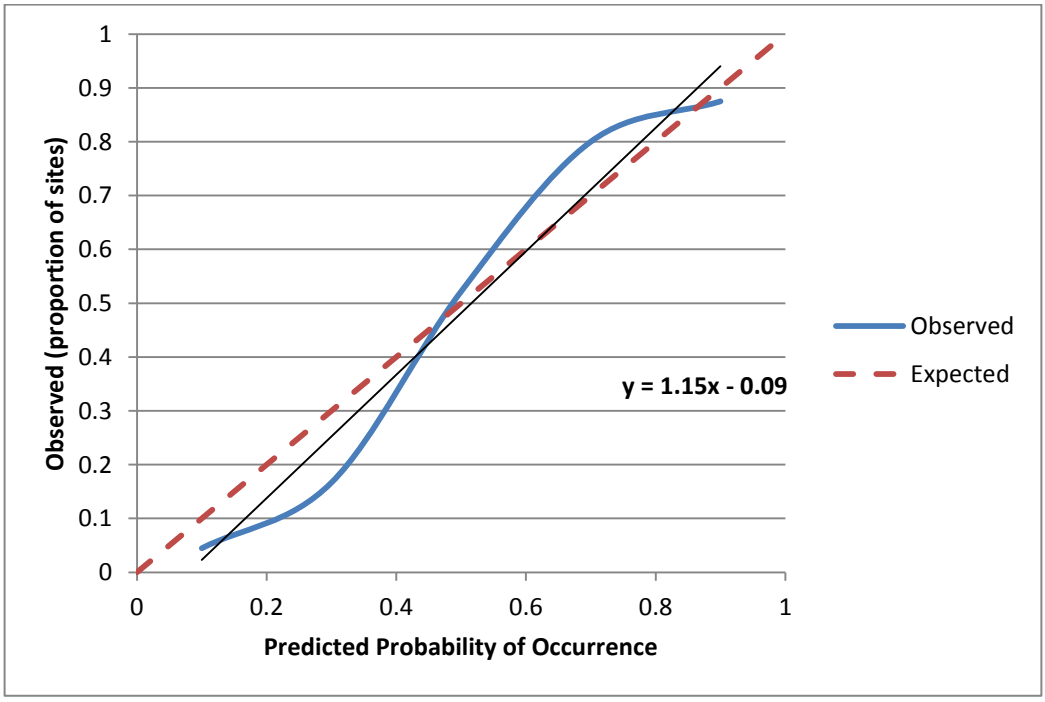

Figure 11. Calibration curve (observed) for the best use-availability model for *D. ocoee* including the expected values and the trendline for the calibration curve.

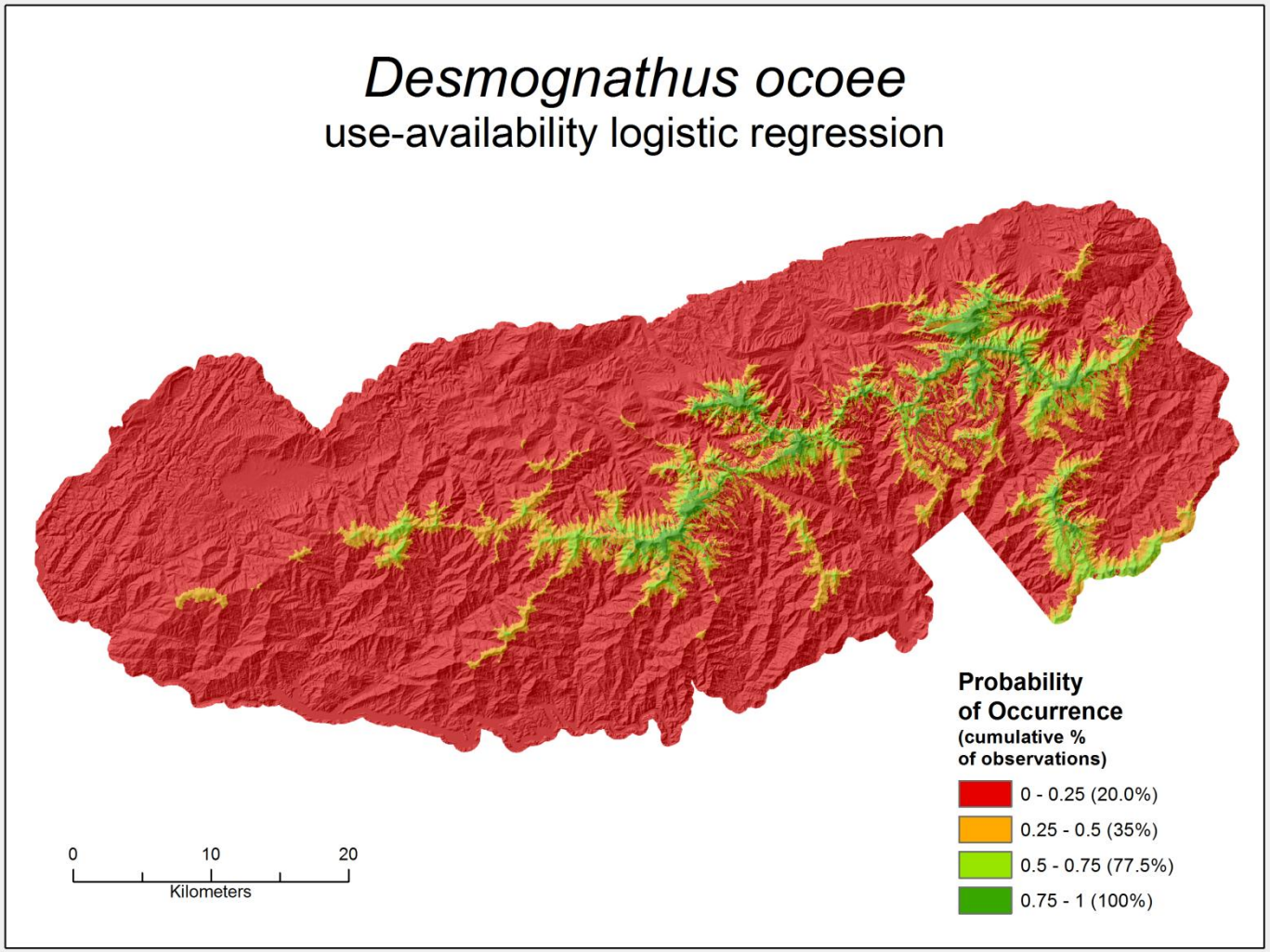

Figure 12. Use-availability logistic regression model extrapolation for *Desmognathus ocoee.*

3.2.2 Case-Control Logistic Regression

The case-control logistic regression models had 73 absence locations that were chosen using other survey sites in the study area where the Ocoee salamander was not found. Of the logistic regression models created, the best model only contained elevation as a parameter (Table 11). In fact, elevation was so dominant that any additional parameters did not improve model performance. Presence of the Ocoee salamander was strongly associated with increasing elevations and the model did not predict locations below approximately 1335 m. Overall, the model predicted 407.3 km^2 or 19.3% of GSMNP was Ocoee salamander habitat. No other univariate model received any support ($\Delta AIC_c > 39.0$).

The best model had a classification accuracy of 85.0% and an area under the ROC (AUC) of 0.94 (Figure 13). That model had a small amount of spread error (slope = 1.13), overestimating from 0.2 to 0.5 but predicted as expected from 0.5 to 1 (Figure 14). Also, the model had some negative bias (y-int = –0.12), skewing the predicted probabilities lower. No other model was available for comparison to the best model. The equation for logistic regression from the best model using only elevation was used to map the model in all of GSMNP (Figure 15).

Table 11. Selection for the case-control logistic regression models for *Desmognathus ocoee*.

Parameters	K	AIC_c	% Cor	R^2	Delta	w
elevation	1	71.1	85.0	0.68		1.0
Breakneck/Oconaluftee soil	1	110.2	81.4	0.40	39.0	0.00
northern hardwood overstory	1	123.6	77.0	0.28	52.5	0.00
undisturbed	1	125.7	76.1	0.26	54.5	0.00
distance to stream	1	130.0	72.6	0.21	58.9	0.00
settled	1	131.5	64.1	0.25	60.4	0.00
heavy disturbance	1	136.7	71.7	0.14	65.6	0.00
spruce-fir overstory	1	138.8	69.9	0.12	67.6	0.00
Spivey soil	1	139.3	64.6	0.11	68.2	0.00
spruce understory	1	139.7	69.9	0.11	68.5	0.00
pine understory	1	139.7	64.1	0.17	68.6	0.00
selective disturbance	1	141.1	64.6	0.09	70.0	0.00
light disturbance	1	144.4	64.6	0.05	73.3	0.00
cove forest overstory	1	144.7	64.6	0.05	73.5	0.00
oak-hickory overstory	1	144.7	64.6	0.05	73.5	0.00
Ditney/Soco soils	1	144.8	64.6	0.05	73.6	0.00
Junaluska soils	1	145.2	64.1	0.11	74.0	0.00

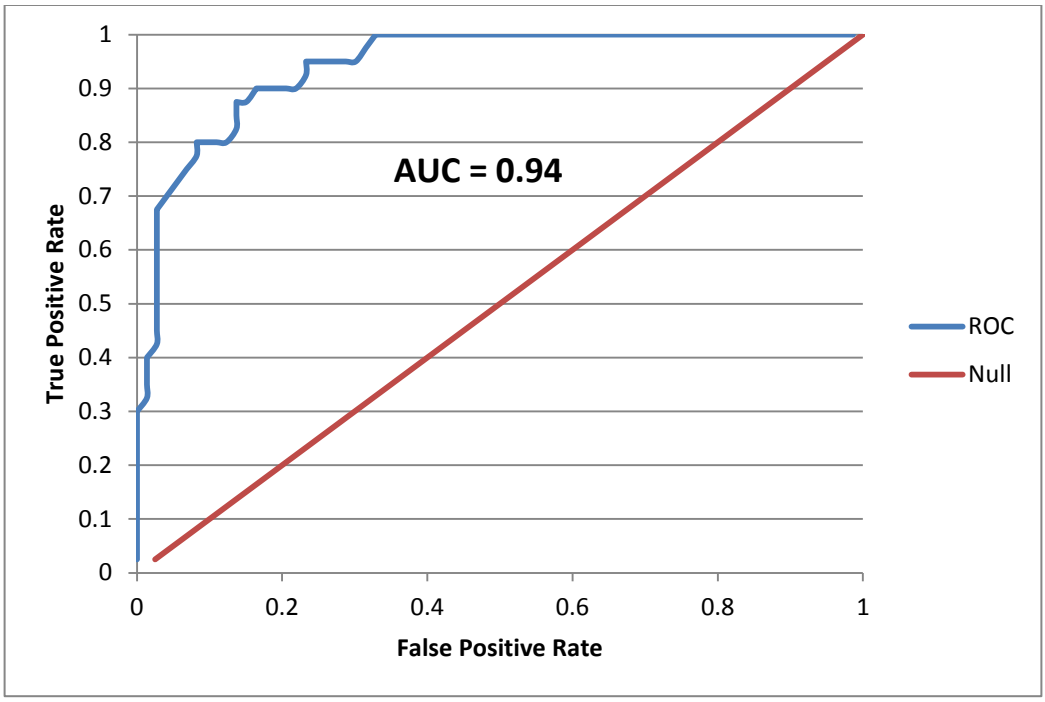

Figure 13. Relative Operating Curve for the best case-control logistic regression model for *D. ocoee.*

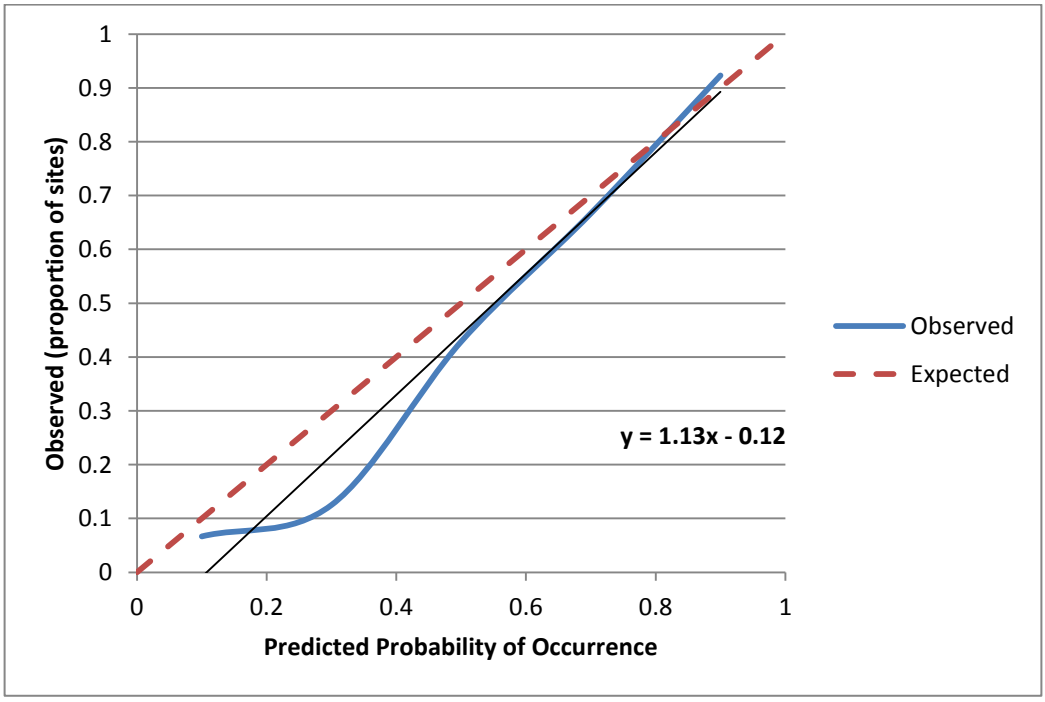

Figure 14. Calibration curve (observed) for the best case-control model for *D. ocoee* including the expected values and the trendline for the calibration curve.

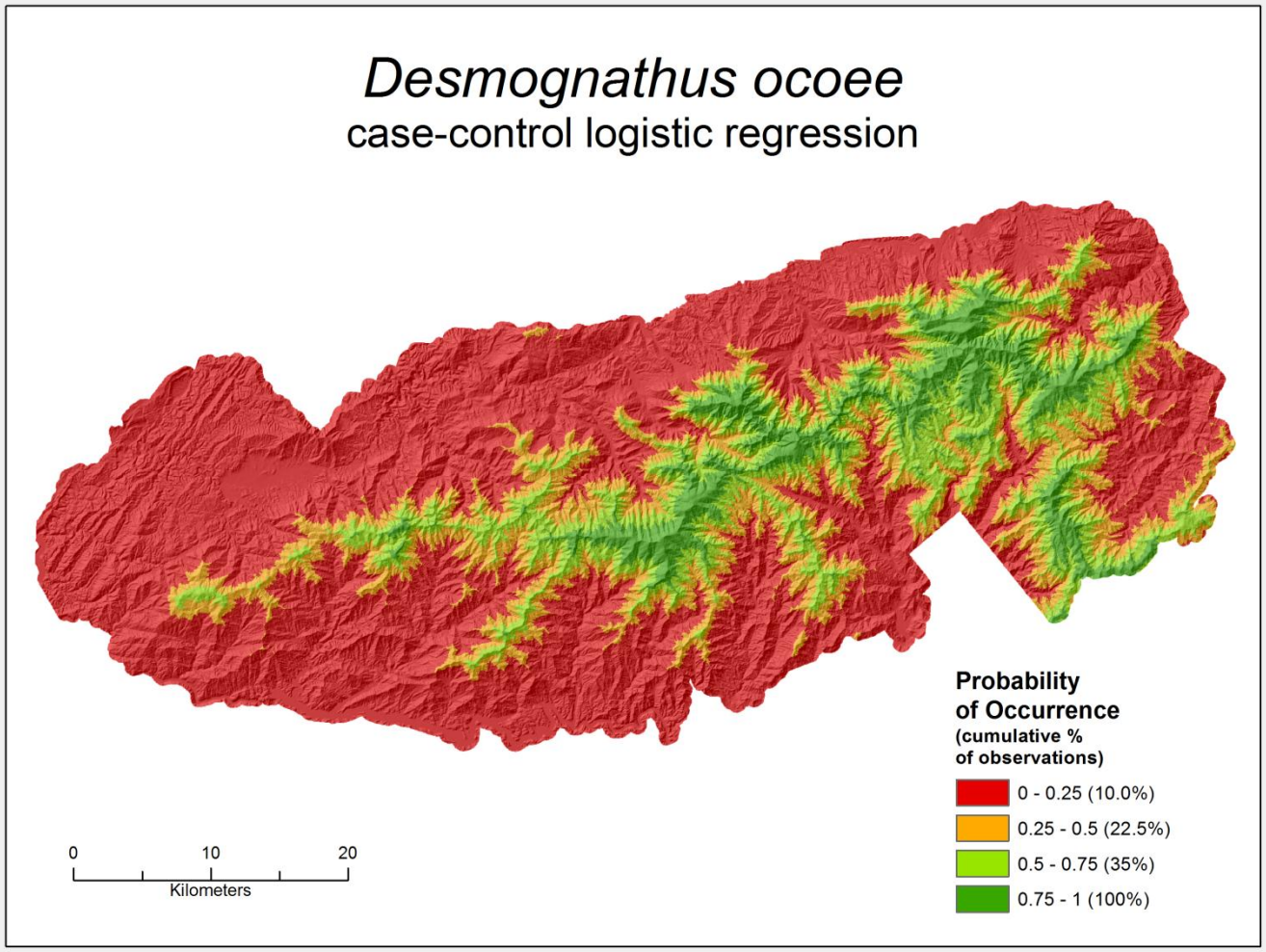

Figure 15. Case-control logistic regression model extrapolation for *Desmognathus ocoee.*

3.2.3 MAHALANOBIS DISTANCE

The data for the Ocoee salamander included 40 unique locations inside GSMNP. The mean values for each parameter are listed in Table 12. Again, I created an original model and then improved on the original. The original Mahalanobis distance values for all of GSMNP ranged from 3.9 to 251.5 ($\bar{x}$ = 53.4, SD = 30.6). The 40 observed Ocoee salamander locations ranged between 7.2 and 26.5 ($\bar{x}$ = 17.0, SD = 5.3), while the 50 random locations had much larger D^2 values ranging between 5.7 and 138.6 ($\bar{x}$ = 46.3, SD = 32.9). The difference between the observed and random locations was significant ($p < 0.01$).

I used the cumulative frequency distributions of the random and observed locations to determine the D^2 value that distinguished habitat favored by the Ocoee salamander from the habitat available in GSMNP. The numerical threshold between favorable and unfavorable habitat, representing the largest distance between the two cumulative frequency graphs, was at a distance value of 26.5 (Figure 16). Ten-fold cross-validation indicated poor model performance with only 37.5% observations remaining under the favorable habitat threshold of 26.5. Accordingly, I removed six variables to improve model performance, leaving the improved model with 11 parameters. I removed variables by identifying equal mean values for continuous variables between favorable and unfavorable habitat. I also removed categorical variables that were represented more in unfavorable than favorable habitat (Table 13).

Table 12. Mean values for variables of *Desmognathus ocoee* locations (* indicates variable was used in final model).

Variable	Mean	Variable	Mean
Elevation*	1511.4	hemlock understory*	0.2
Slope	21.3	herbaceous understory*	0.6
Beers	1.1	*Rhododendron* understory*	0.4
RSP	56.3	spruce understory*	0.2
TRMI	28.7	high-elevation beech overstory	0.1
historically heavily disturbed*	0.3	northern hardwood overstory*	0.6
historically undisturbed*	0.6	spruce-fir overstory*	0.2
Breakneck-Oconaluftee soils*	0.6	distance to stream*	285.0
Ditney-Soco soils	0.1		

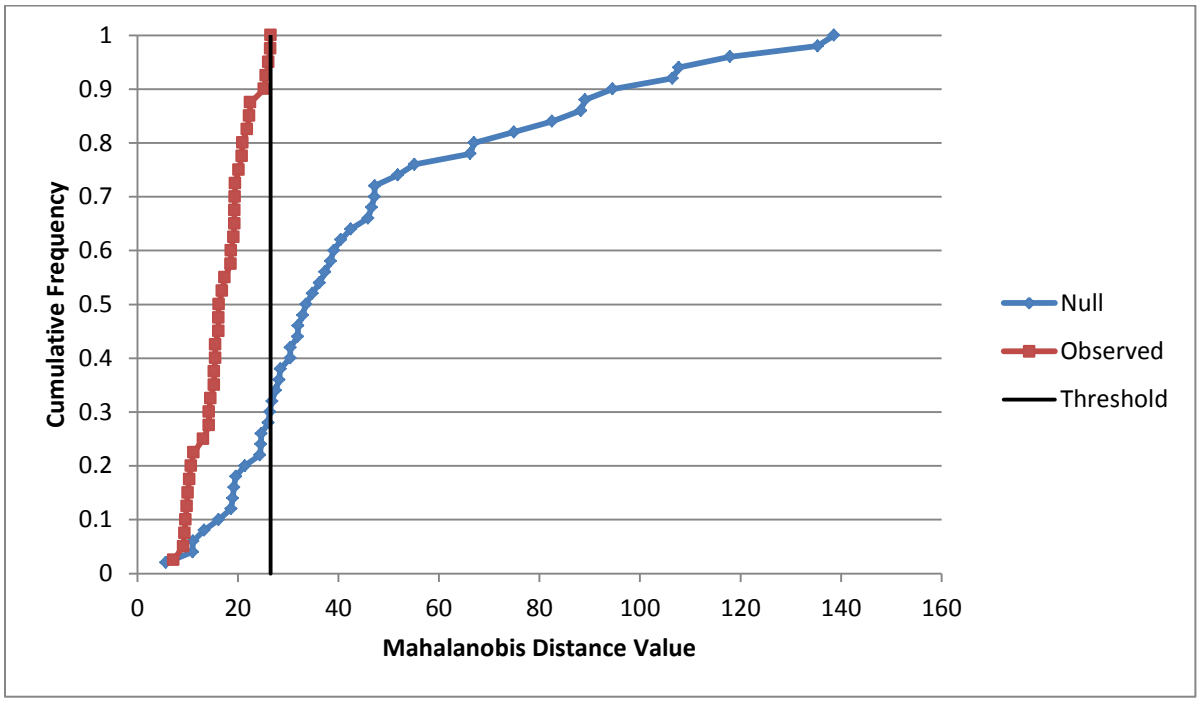

Figure 16. Cumulative frequency graph for the original Mahalanobis distance model versus the random (null) model for *Desmognathus conanti.*

Table 13. Favorable versus unfavorable habitat based on the original model for *D. ocoee*. Continuous variables are compared by mean and categorical variables are compared by their proportion (* indicates variable not used in final model).

	Favorable		Unfavorable	
Variable	**Mean**	**SD**	**Mean**	**SD**
Elevation*	1370.0	221.9	939.0	317.6
Slope	24.7	7.1	22.4	9.5
Beers	1.0	0.7	1.0	0.7
RSP	52.1	40.1	49.9	42.2
TRMI	27.1	9.8	27.9	11.3
Distance to stream*	216.6	148.1	185.3	152.3
	Prop		**Prop**	
Historically heavily disturbed*	0.310		0.157	
Historically undisturbed*	0.563		0.159	
Breakneck-Oconaluftee soils*	0.445		0.082	
Ditney-Soco soils	0.319		0.515	
Hemlock understory*	0.200		0.142	
Herbaceous understory*	0.541		0.453	
Rhododendron understory*	0.455		0.344	
Spruce understory*	0.142		0.026	
High-elevation beech overstory	0.018		0.013	
Northern hardwood overstory*	0.553		0.145	
Spruce-fir overstory*	0.080		0.017	

The improved model contained 11 of the original 17 parameters. The D^2 values for all of GSMNP ranged between 3.6 and 110.9 ($\bar{x}$ = 29.8, SD = 19.5). The 40 Ocoee salamander locations ranged between 3.8 and 22.8 ($\bar{x}$ = 11.0, SD = 5.0). Conversely, the random locations had distances between 3.8 and 73.6 ($\bar{x}$ = 30.2, SD = 20.2), and the means of the observed and random locations were significantly different ($p < 0.01$). Using the cumulative frequency graph, I determined the habitat threshold was at a D^2 value of 17.5 (Figure 17). Overall, 95% of the 40 Ocoee salamander locations had D^2 values below 17.5. Other thresholds used to more easily visualize Ocoee salamander habitat were $D^2 < 6.8$ (25%), $D^2 < 10.5$ (50%), $D^2 < 17.5$ (95%), $D^2 < 22.8$ (100%). Pixels with values greater than 22.8 constituted the sixth class.

Next I calculated D^2 values for all pixels in GSMNP. I estimated 746 km^2 of 2,108 km^2 of GSMNP was favorable habitat for the Ocoee salamander (Figure 18). The favorable habitat was concentrated in the high elevations through the middle and eastern portions of the park. Lower elevations on the western side were deemed unfavorable. Elevation and nearness to streams were the two most important habitat characteristics. The means of the locations for both variables were significantly different than random locations ($p < 0.01$), the coefficients of variation for observed locations were smaller random locations and smaller than other variables (Table 14), and both variables showed strong correlations with other variables in the model. Elevation in favorable habitat was higher ($\bar{x}$ = 1306.7, SD = 210.3) than in unfavorable habitat ($\bar{x}$ = 852.0, SD = 293.2). Also, favorable habitat was farther from streams ($\bar{x}$ = 204.3, SD = 148.7) than unfavorable habitat ($\bar{x}$ = 183.3, SD = 153.3). Similar to *D. conanti*, *D. ocoee* was farther from streams as elevation increased (ρ = 0.38, $p < 0.02$).

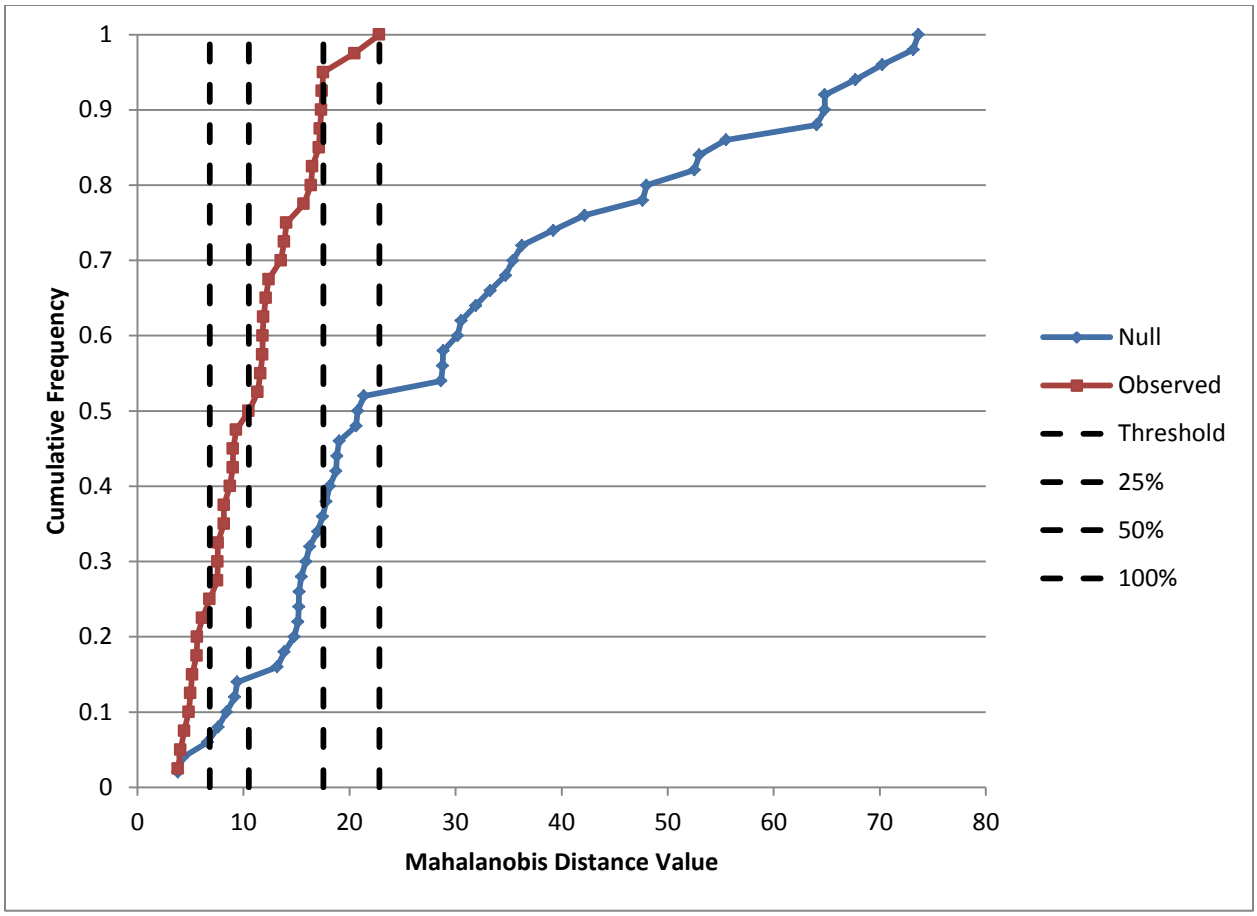

Figure 17. Cumulative frequency graph for the improved Mahalanobis distance model versus the random (null) model for *Desmognathus ocoee.*The solid line is the threshold between good and bad habitat. The dashed lines represent the cumulative percentage of salamander observations below each line.

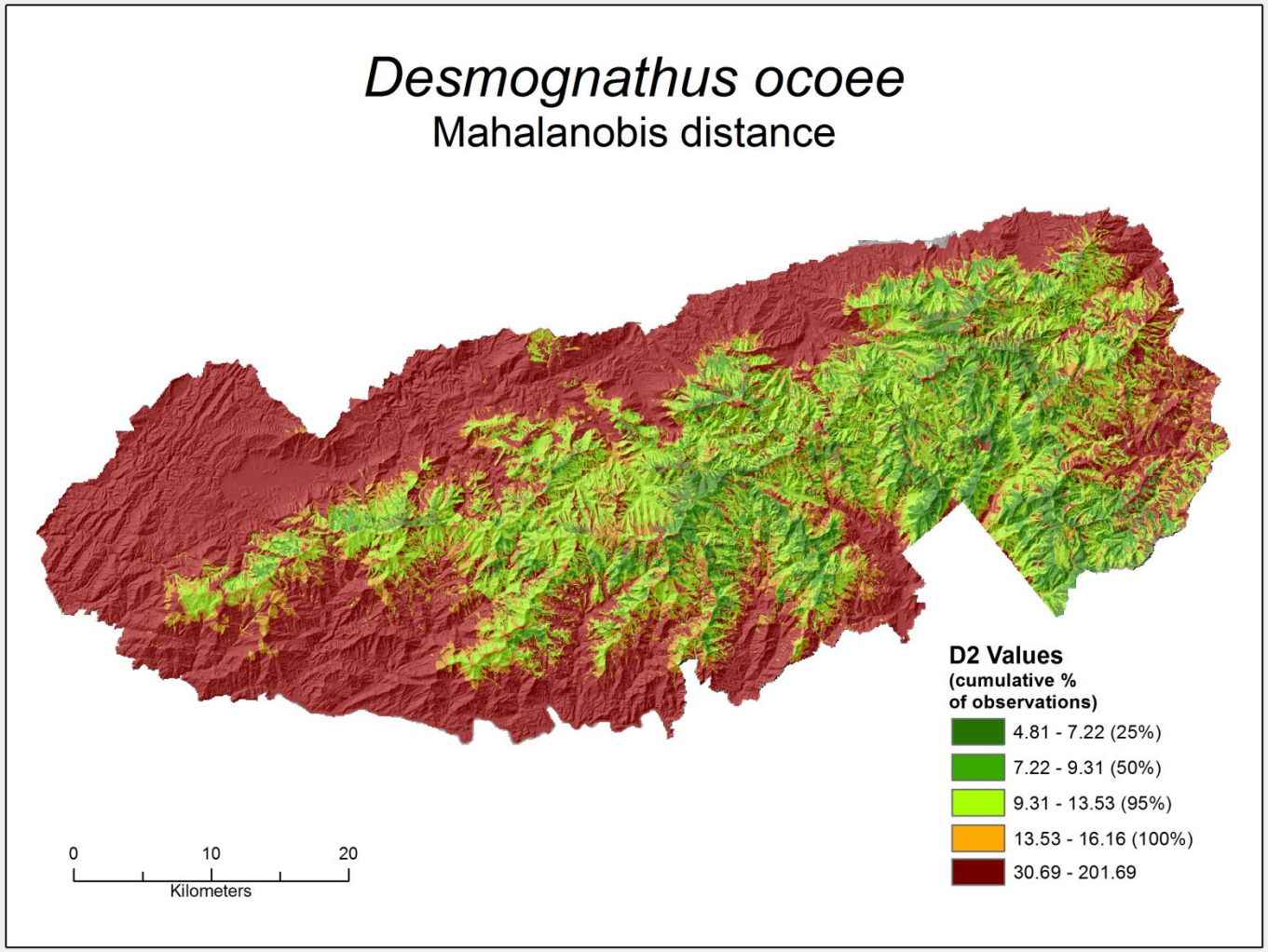

Figure 18. Mahalanobis distance model for *Desmognathus ocoee.*

Table 14. Coefficients of variation for favorable habitat compared to GSMNP as a whole for each variable used in the *D. ocoee* Mahalanobis distance model.

Variable	Favorable	GSMNP
Elevation	0.32	0.35
Historically heavily disturbed	1.46	2.17
Historically undisturbed	0.92	1.83
Breakneck/Oconaluftee soil	0.87	2.50
Hemlock understory	2.41	2.33
Herbaceous understory	0.92	1.06
Rhododendron understory	1.38	1.37
Spruce understory	2.03	5.00
Northern hardwood overstory	0.92	2.00
Spruce-fir overstory	2.20	5.33
Distance to streams	0.70	0.80

Breakneck/Oconaluftee soils, northern hardwood overstory, and historically undisturbed areas seemed to be very important as well. Only 6.3% of the unfavorable area was comprised of Breakneck-Oconaluftee soils, but 29.2% of predicted favorable habitat was comprised of those soils. Breakneck/Oconaluftee soils are both frigid sandstone and their cool temperatures help retain more water. The vegetation cover in the soils is generally spruce-fir (Breakneck) and northern hardwood (Oconaluftee). For the Ocoee salamander, occurrences are more likely in these soils at relatively higher elevations (r_{pb} = 0.54, $p < 0.01$). Similarly, only 7.5% of unfavorable areas contained northern hardwood overstory, but that overstory type was in 47.0% of the favorable habitat. Northern hardwood overstory was not significantly correlated with any other non-overstory variable among the 40 Ocoee salamander locations. Historically undisturbed areas were also not correlated with any non-disturbance variable, but were represented at 49.7% in favorable habitat versus only 8.0% in unfavorable areas.

Other variables included in the model were heavily disturbed areas; the understories: hemlock, herbaceous, *Rhododendron*, and spruce; and spruce-fir overstory. Heavily disturbed areas comprised 32.0% of favorable habitat versus unfavorable and was not correlated with any non-disturbance variable. The Ocoee salamander was more likely in hemlock understory at relatively lower elevations ($r_{pb} = -0.53$, $p < 0.01$) and spruce understory at higher elevations ($r_{pb} = 0.38$, $p < 0.01$). Ocoee salamanders were also less likely to be in hemlock understory if the soil was Breakneck-Oconaluftee types ($\phi = -0.35$, $p < 0.03$). Favorable habitat was 24.7% hemlock and 10.4% spruce while unfavorable areas were 10.1% hemlock and 1.4% spruce. Herbaceous understory, *Rhododendron* understory, and spruce-fir overstory composed 50.4%, 49.3%, and 5.5% of favorable habitat, and only

44.7%, 29.2%, and 1.2% of unfavorable areas, respectively. None of these were correlated significantly with variables not of the same general category.

The improved model showed a large improvement over the original model in the ten-fold cross-validation. Across the ten iterations, the percentage of observations that stayed under the favorable habitat threshold of 17.54 ranged between 25 and 100% with an overall average of 60.0%. This is compared to the low percentage of 37.5% in the original model. The improved model indicates a better performance than the original.

3.3 *PLETHODON JORDANI*

3.3.1 USE-AVAILABILITY LOGISTIC REGRESSION

Of the logistic regression models created, the best model contained the variables elevation and spruce understory (Table 15). The weight of evidence (w_i = 0.79) for the best model was 3.7 times greater than the second best model. Presence of the red-cheeked salamander was positively associated with higher elevation, occurring above 1300 m, but negatively associated with spruce understory. Elevation alone was included in the second best model and was nearly considered the best model (ΔAIC_c = 2.6; w_i = 0.21). This model again indicated the salamander was positively associated with increasing elevation. The remaining univariate models did not receive any support (ΔAIC_c > 68.3).

The best model had a classification accuracy of 73.6% and an area under the ROC (AUC) of 0.83 (Figure 19). It had a small amount of spread error (slope = 1.01), over- and underestimating by small amounts that balanced over the whole model (Figure 20). The model also had a negligible amount of bias (y-int < –0.01). The second-best model had an accuracy of 74.7% and an AUC of 0.82. It had a slightly larger spread error (slope = 0.97) and bias (y-int < 0.01) than the best model. The equation for logistic regression from the best model was used to map the model across the landscape of GSMNP (Figure 21). Once mapped I determined that 353 km^2 was considered red-cheeked salamander habitat.

Table 15. Selection for the use-availability logistic regression models for *Plethodon jordani*.

Parameters	K	AIC_c	% Cor	R^2	Delta	w
elevation, spruce understory	2	282.2	73.6	0.41		0.79
elevation	1	284.9	75.1	0.39	2.6	0.21
Breakneck/Oconaluftee soil	1	350.5	65.2	0.14	68.3	0.00
undisturbed	1	360.0	63.8	0.09	77.8	0.00
distance to stream	1	363.1	59.3	0.08	80.9	0.00
northern hardwood overstory	1	363.4	61.9	0.08	81.2	0.00
Ditney/Soco soil	1	365.2	60.4	0.07	82.9	0.00
Luftee soil	1	368.3	57.1	0.06	86.1	0.00
spruce-fir overstory	1	369.2	57.1	0.05	87.0	0.00
pine overstory	1	371.5	52.0	0.04	89.2	0.00
spruce understory	1	371.9	57.5	0.04	89.7	0.00
oak-hickory overstory	1	373.0	53.8	0.03	90.8	0.00
selective disturbance	1	374.7	52.4	0.03	92.5	0.00
cove forest overstory	1	375.9	53.1	0.02	93.7	0.00
chestnut oak overstory	1	376.2	52.0	0.02	93.9	0.00
light disturbance	1	376.3	53.8	0.02	94.1	0.00

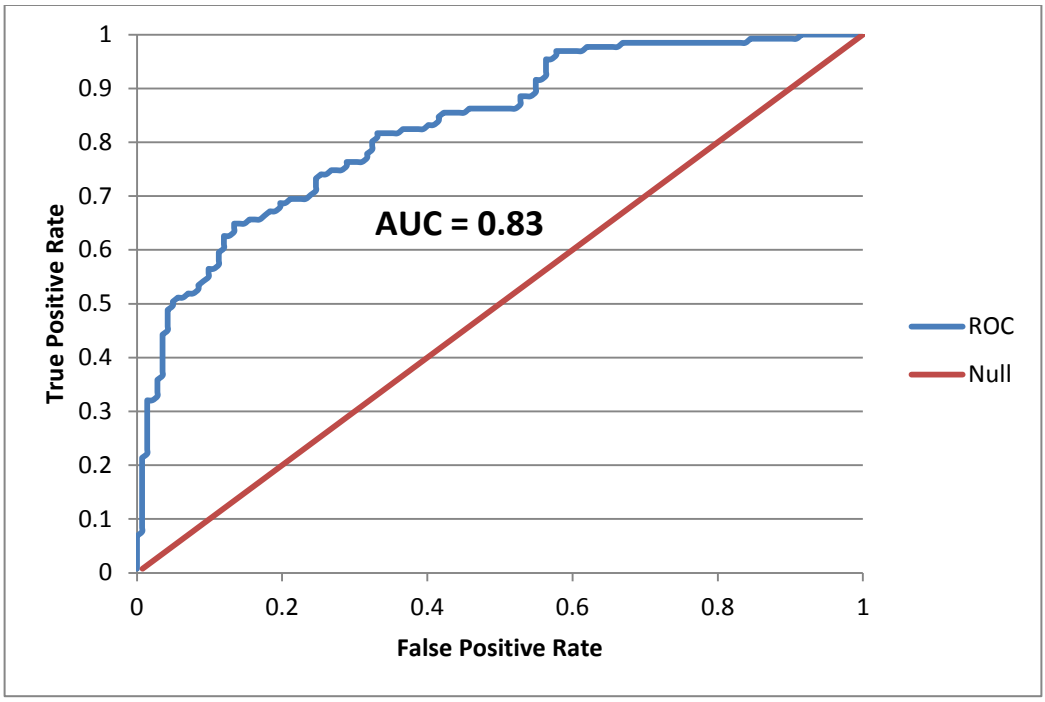

Figure 19. Relative Operating Curve for the best use-availability logistic regression model for *P. jordani*.

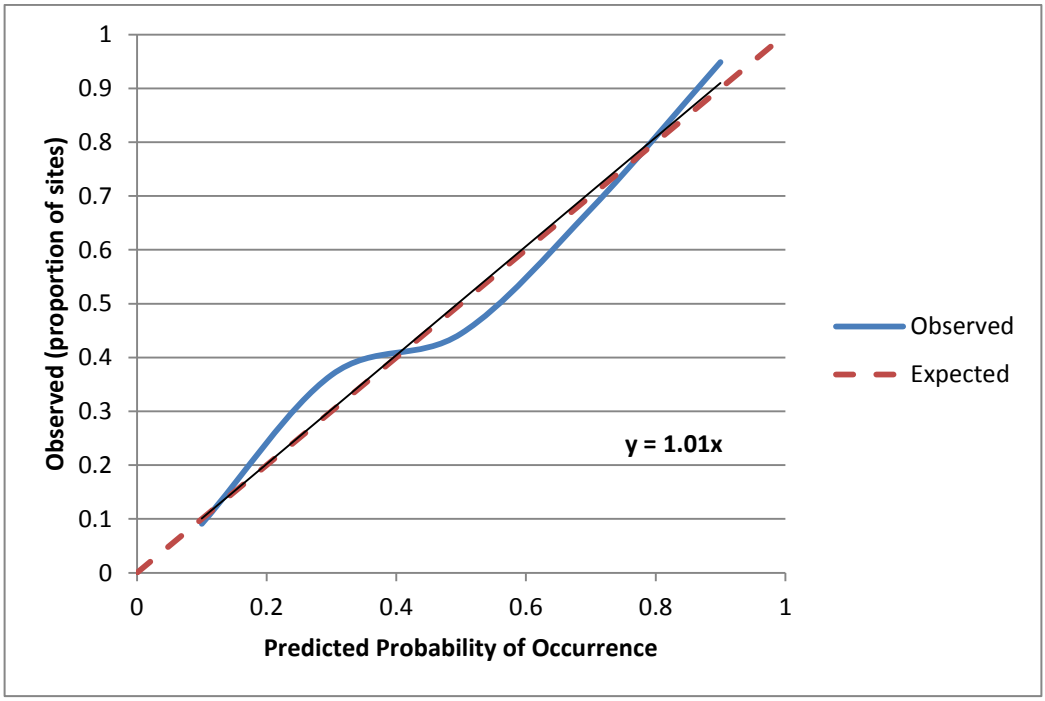

Figure 20. Calibration curve (observed) for the best use-availability model for *P. jordani* including the expected values and the trendline for the calibration curve.

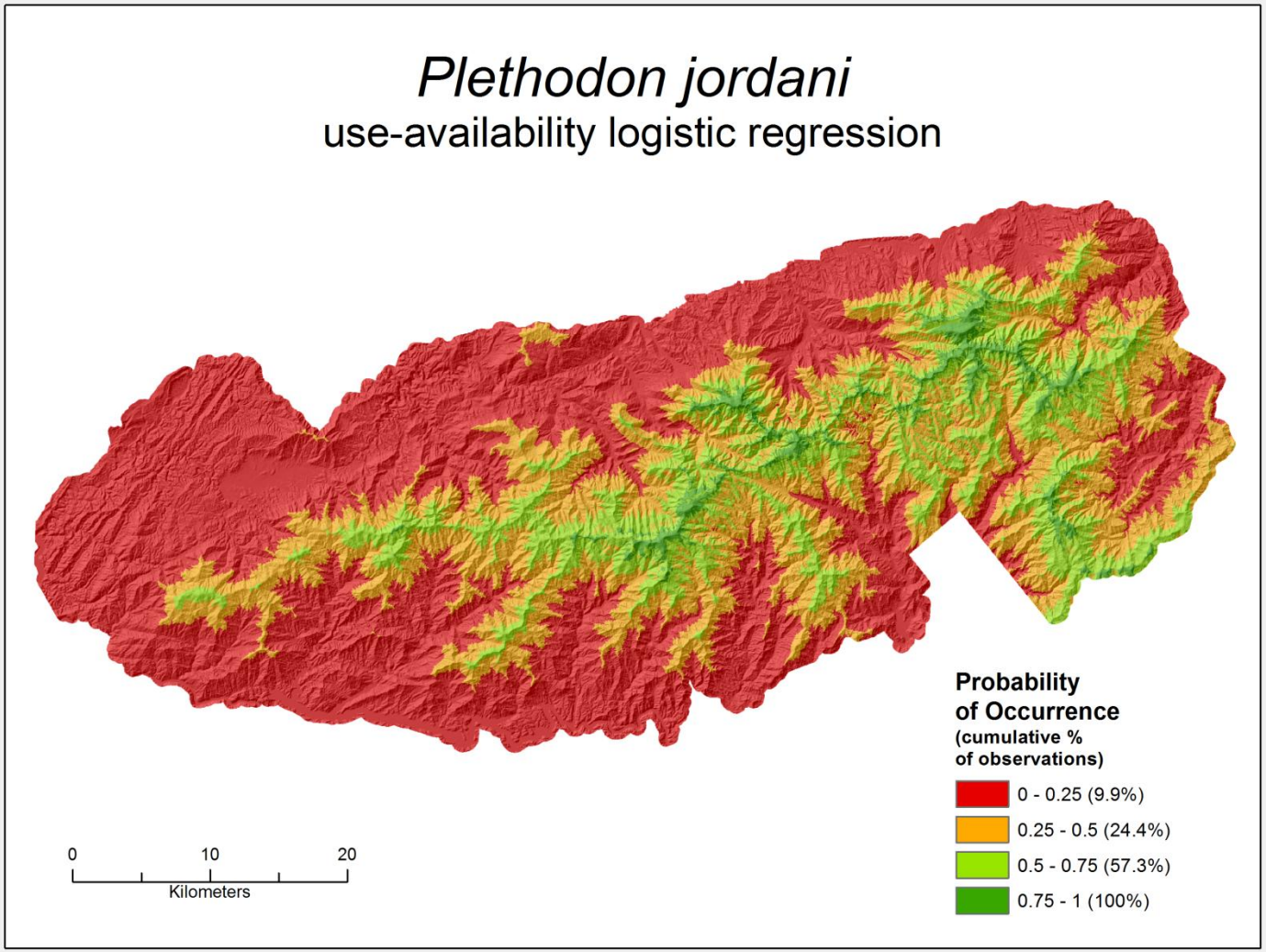

Figure 21. Use-availability logistic regression model extrapolation for *Plethodon jordani*.

3.3.2 Case-Control Logistic Regression

The absence locations in this model were generated using other survey sites in the study area. Of the case-control logistic regression models created, the best model included elevation, pine understory, northern hardwood overstory, and distance to streams (Table 16). The weight of evidence (w_i = 0.84) for the best model was over five times greater than the second best model. Presence of the red-cheeked salamander was positively associated only with higher elevations and was predicted in areas above 1000 m. The salamander was negatively associated with pine understory, northern hardwood overstory, and areas farther from streams. The next best model, "elevation, northern hardwood overstory, and proximity to streams", received some support (ΔAIC_c = 3.3; w_i = 0.16). This model showed the salamander was positively associated with increasing elevation but negatively associated with northern hardwood overstory and increasing distance from streams. All other models did not receive any support (ΔAIC_c > 10.3).

The best model had a classification accuracy of 84.2% and an area under the ROC (AUC) of 0.92 (Figure 22). The best model had a small amount of spread error (slope = 1.02), underestimating slightly more than overestimating from 0 to 1 (Figure 23). The model had a small amount of positive bias (y-int = 0.03), skewing the predicted probabilities higher. The second-best model also had an accuracy of 84.2% and an AUC of 0.916. It also had a tiny amount of spread error (slope = 1.04) and a very small amount of bias (y-int = 0.01). The model underestimates very slightly around 0.3, but mostly predicts as expected. The equation for logistic regression from the best model was used to calculate probability of red-cheeked salamander occurrence in all areas of GSMNP (Figure 24). Overall, 618 km^2 of GSMNP was predicted to be *P. jordani* habitat.

Table 16. Selection for the case-control logistic regression models for *Plethodon jordani*.

Parameters	K	AIC_c	% Cor	R^2	Delta	w
elevation, pine understory, northern hardwood overstory, distance to stream	4	206.7	84.2	0.65	0	0.83
elevation, northern hardwood overstory, distance to stream	3	210.0	84.2	0.64	3.3	0.16
elevation	1	217.0	83.1	0.61	10.2	0.00
undisturbed	1	331.4	70.5	0.24	124.7	0.00
Breakneck/Oconaluftee soil	1	334.2	69.1	0.23	127.5	0.00
pine understory	1	345.3	59.4	0.18	138.6	0.00
spruce understory	1	348.4	62.6	0.17	141.7	0.00
distance to stream	1	352.0	70.5	0.16	145.3	0.00
northern hardwood overstory	1	356.2	65.5	0.14	149.5	0.00
spruce-fir overstory	1	363.1	59.4	0.11	156.4	0.00
pine overstory	1	369.1	54.0	0.08	162.3	0.00
oak-hickory overstory	1	370.4	56.8	0.08	163.6	0.00
settled	1	371.4	57.9	0.07	164.6	0.00
selective disturbance	1	378.4	52.9	0.04	171.7	0.00
herbaceous understory	1	379.2	58.3	0.04	172.4	0.00
slope	1	379.8	55.4	0.03	173.1	0.00
light disturbance	1	380.8	54.7	0.03	174.1	0.00
chestnut oak overstory	1	382.3	52.9	0.02	175.6	0.00
Ditney/Soco soil	1	382.4	54.3	0.02	175.7	0.00

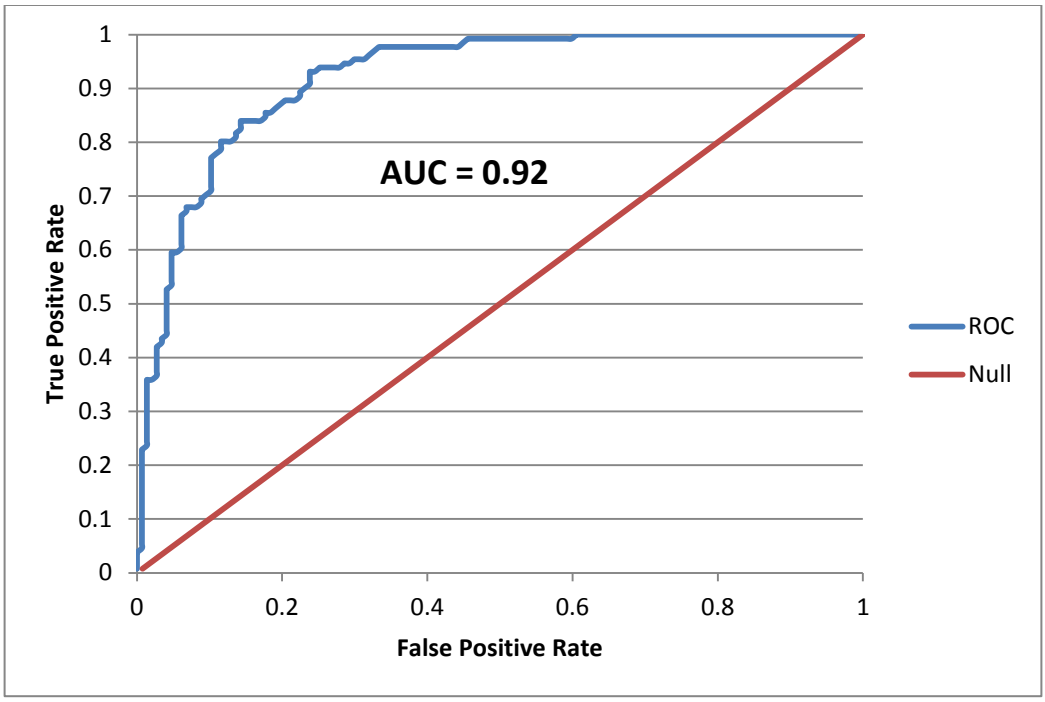

Figure 22. Relative Operating Curve for the best case-control model for *P. jordani*.

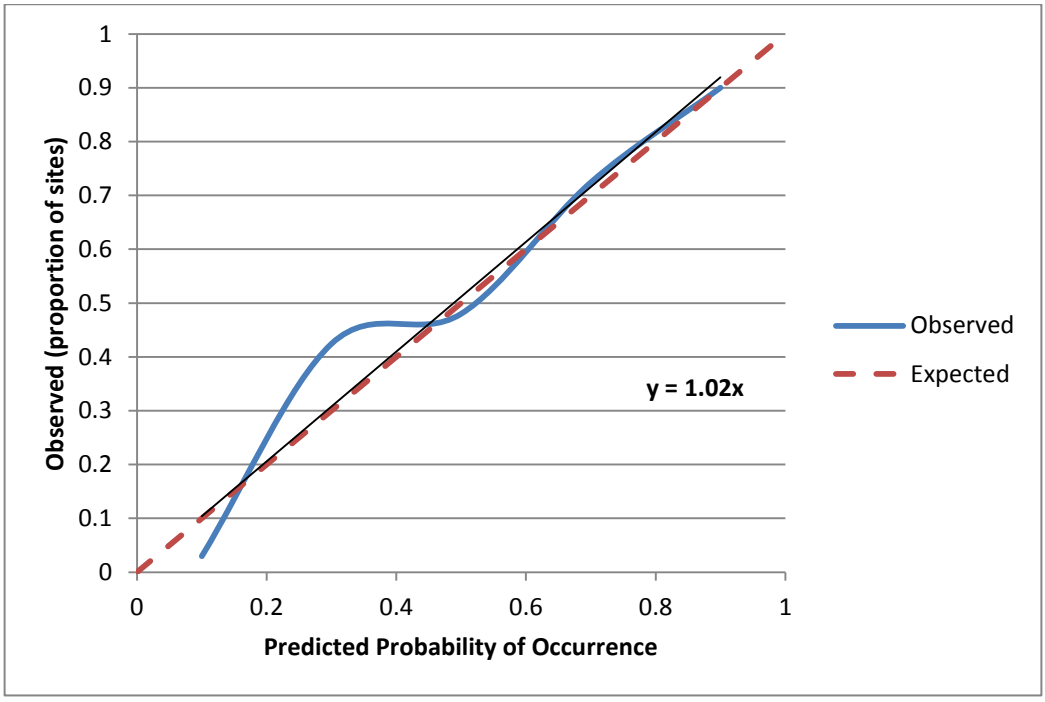

Figure 23. Calibration curve (observed) for the best case-control model for *P. jordani* including the expected values and the trendline for the calibration curve.

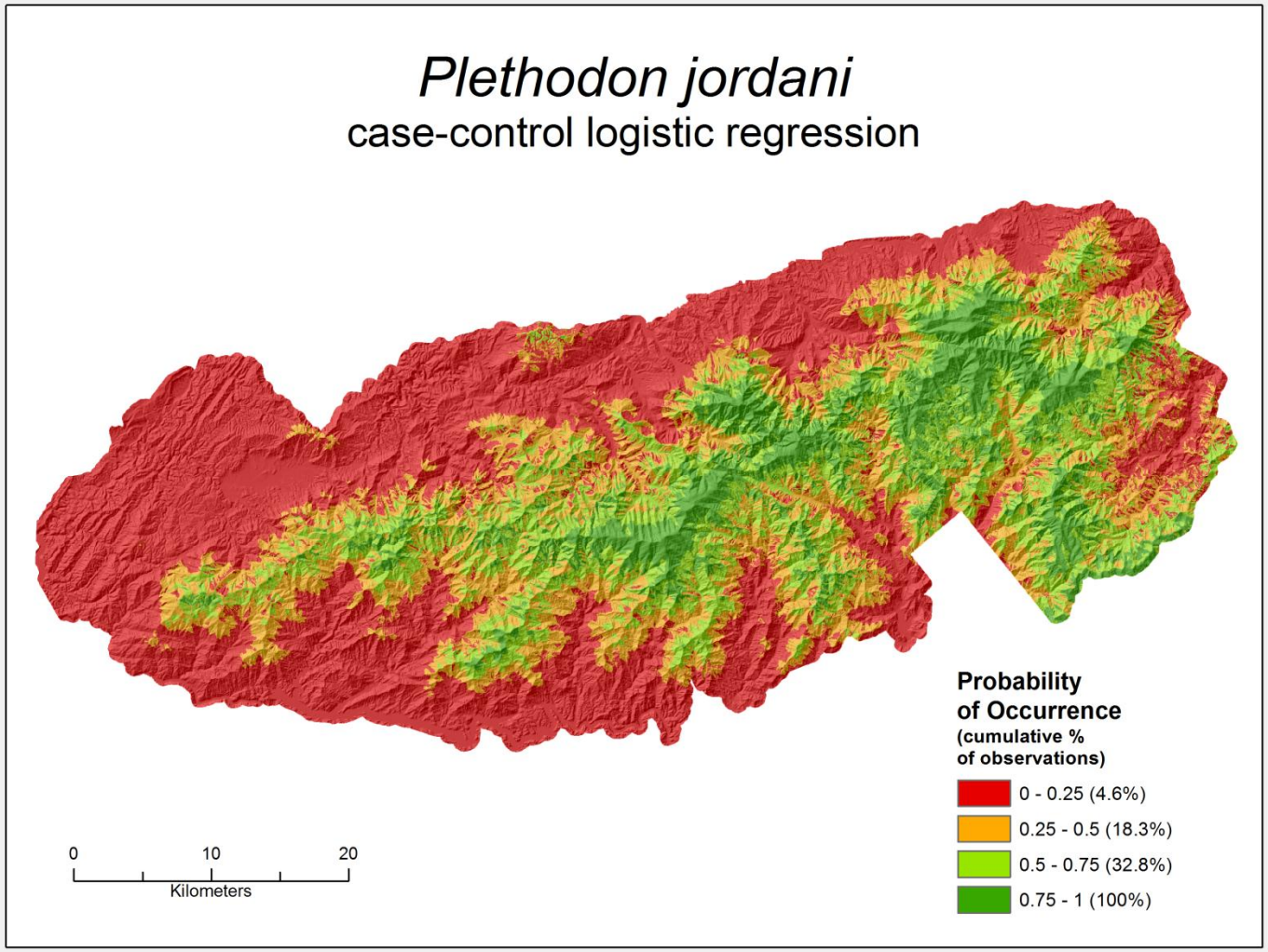

Figure 24. Case-control logistic regression model extrapolation for *Plethodon jordani*.

3.3.3 MAHALANOBIS DISTANCE

The red-cheeked salamander had 131 unique locations from the dataset inside GSMNP. The mean values for each parameter are listed in Table 17. The first model that was later refined had Mahalanobis distance values between 4.5 and 82.6 ($\bar{x}$ = 21.9, SD = 7.7). The 131 observed salamander locations ranged from 4.8 to 31.7 ($\bar{x}$ = 16.0, SD = 5.0), whereas the 131 random (null model) locations had values that ranged from 8.7 to 51.6 ($\bar{x}$ = 22.7, SD = 7.7). The difference between the means of the observed and random locations was significant ($p < 0.01$).

I was able to determine habitat used by red-cheeked salamander and separate it from unused areas of GSMNP by using the cumulative frequency distributions of the distance values of the observed locations and random locations. The point on the graph representing the greatest distance between the two curves, and therefore the threshold between favorable and unfavorable habitat was at the D^2 value of 18.4 (Figure 25). A ten-fold cross-validation on the original model resulted in only 52.7% of the 131 observations remaining under the 18.4 threshold. To improve the model, I removed eight parameters that either did not differ in mean values between favorable and unfavorable habitat or were underrepresented in favorable versus unfavorable habitat (Table 18).

Table 17. Mean values for variables of *Plethodon jordani* locations (* indicates variable was used in final model).

Variable	Mean	Variable	Mean
Elevation*	1427.4	Ditney-Soco soils*	0.2
Slope	22.8	Luftee soils	0.1
RSP	49.4	herbaceous understory*	0.6
TRMI	28.6	spruce understory*	0.2
historically heavily disturbed	0.2	cove overstory	0.1
historically lightly disturbed	0.2	northern hardwood overstory*	0.4
historically undisturbed*	0.5	spruce-fir overstory	0.1
Breakneck-Oconaluftee soils*	0.4	distance to stream*	267.0

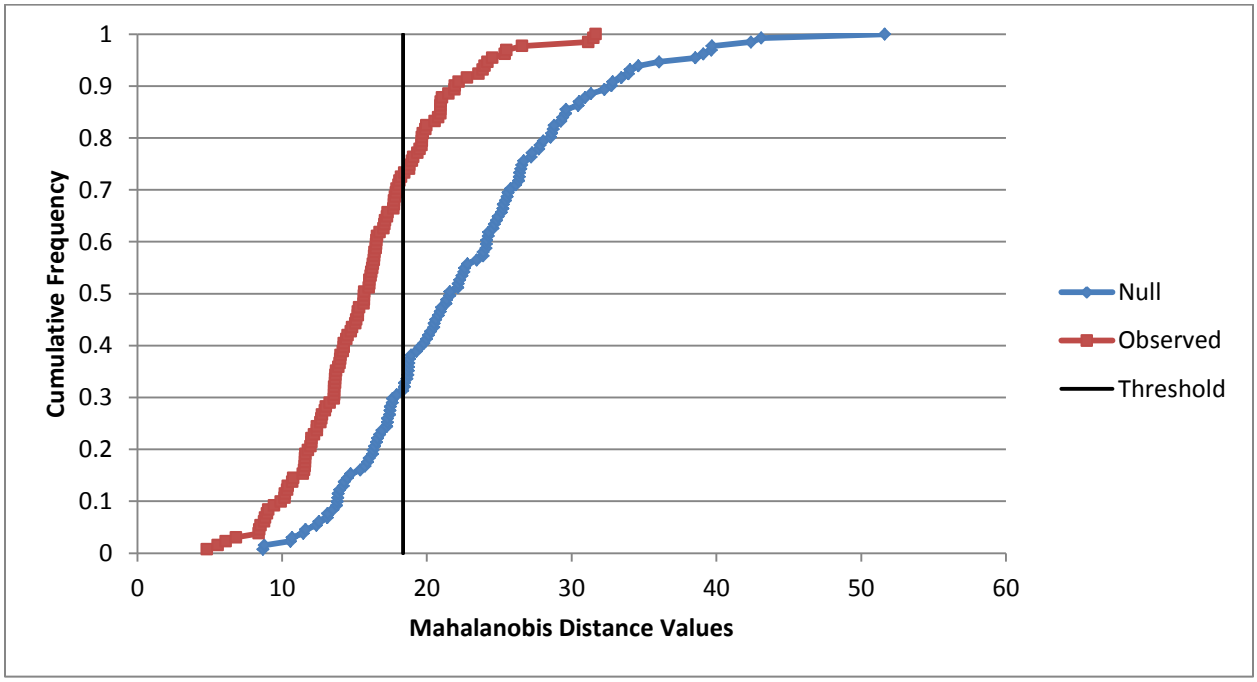

Figure 25. Cumulative frequency graph for the original Mahalanobis distance model versus the random (null) model for *Plethodon jordani.*

Table 18. Favorable versus unfavorable habitat based on the original model for *P. jordani*. Continuous variables are compared by mean and categorical variables are compared by their proportion (* indicates variable used in final model).

	Favorable		Unfavorable	
Variable	**Mean**	**SD**	**Mean**	**SD**
Elevation*	1120.7	255.0	895.0	332.1
Slope	24.1	7.0	22.1	10.1
RSP	52.5	39.9	49.0	42.9
TRMI	26.9	9.9	28.2	11.7
Distance to stream*	214.1	139.3	177.6	157.4
	Prop		**Prop**	
Historically heavily disturbed	0.156		0.199	
Historically lightly disturbed	0.230		0.322	
Historically undisturbed*	0.383		0.140	
Breakneck-Oconaluftee soils*	0.271		0.073	
Ditney-Soco soils*	0.545		0.446	
Luftee soils	0.028		0.038	
Herbaceous understory*	0.542		0.425	
Spruce understory*	0.081		0.026	
Cove overstory	0.070		0.220	
Northern hardwood overstory*	0.292		0.171	
Spruce-fir overstory	0.033		0.024	

By removing half of the parameters from the model, I was able to increase the performance of the model. The ten-fold cross-validation showed that 75.6% Mahalanobis values stayed under the new threshold of 10.29 across the ten trials. The distance values for GSMNP ranged from 4.1 to 55.7 ($\bar{x}$ = 10.4, SD = 4.6) while the values for the observed red-cheeked salamander location were more contained between 4.3 and 18.2 ($\bar{x}$ = 8.0, SD = 2.6). The distances of the random locations were overall larger, ranging between 4.2 and 26.5 ($\bar{x}$ = 10.5, SD = 4.5) and were significantly different from observed locations ($p < 0.01$). I used the greatest distance between the cumulative frequency distributions of the observed and random locations to determine a favorable habitat threshold of 10.3 (Figure 26). Of all 131 red-cheeked salamander locations, 86.3% of the locations had D^2 values below 10.29. Other thresholds used to more easily visualize red-cheeked salamander habitat were $D^2 < 6.0$ (25%), $D^2 < 7.7$ (50%), $D^2 < 10.3$ (74.8%), $D^2 < 12.7$ (95%), $D^2 < 18.2$ (100%). Pixels with values greater than 18.2, areas with no observations, constituted the sixth class.

After calculating Mahalanobis distance for all of GSMNP, I estimated 1,240 km^2 of the 2,108 km^2 of GSMNP was red-cheeked salamander habitat (Figure 27). The favorable habitat was mostly continuous above elevations around 700 m. The lowest elevations on the western side were deemed unfavorable. Elevation and nearness to streams were the two habitat characteristics representing continuous variables and were strong variables as usual. They were both significantly different than values available in GSMNP ($p < 0.01$) and had lower coefficients of variation than other variables in the model (Table 19). Elevation in favorable habitat was higher ($\bar{x}$ = 1143.8, SD = 270.4) than in unfavorable habitat ($\bar{x}$ = 826.4, SD = 351.8). In the improved model, there was little difference between the distance

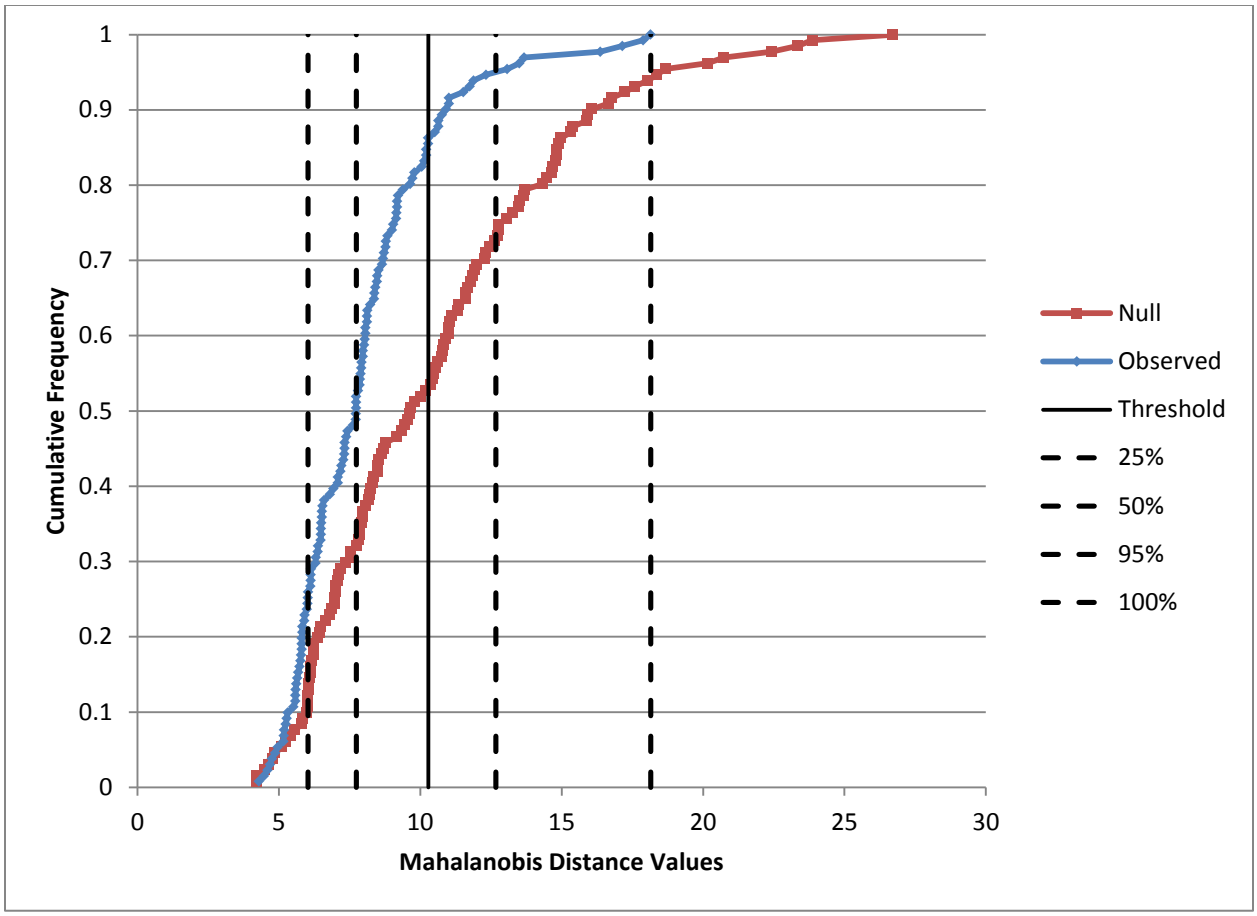

Figure 26. Cumulative frequency graph for the improved Mahalanobis distance model versus the random (null) model for *Plethodon jordani.*The solid line is the threshold between good and bad habitat. The dashed lines represent the cumulative percentage of salamander observations below each line.

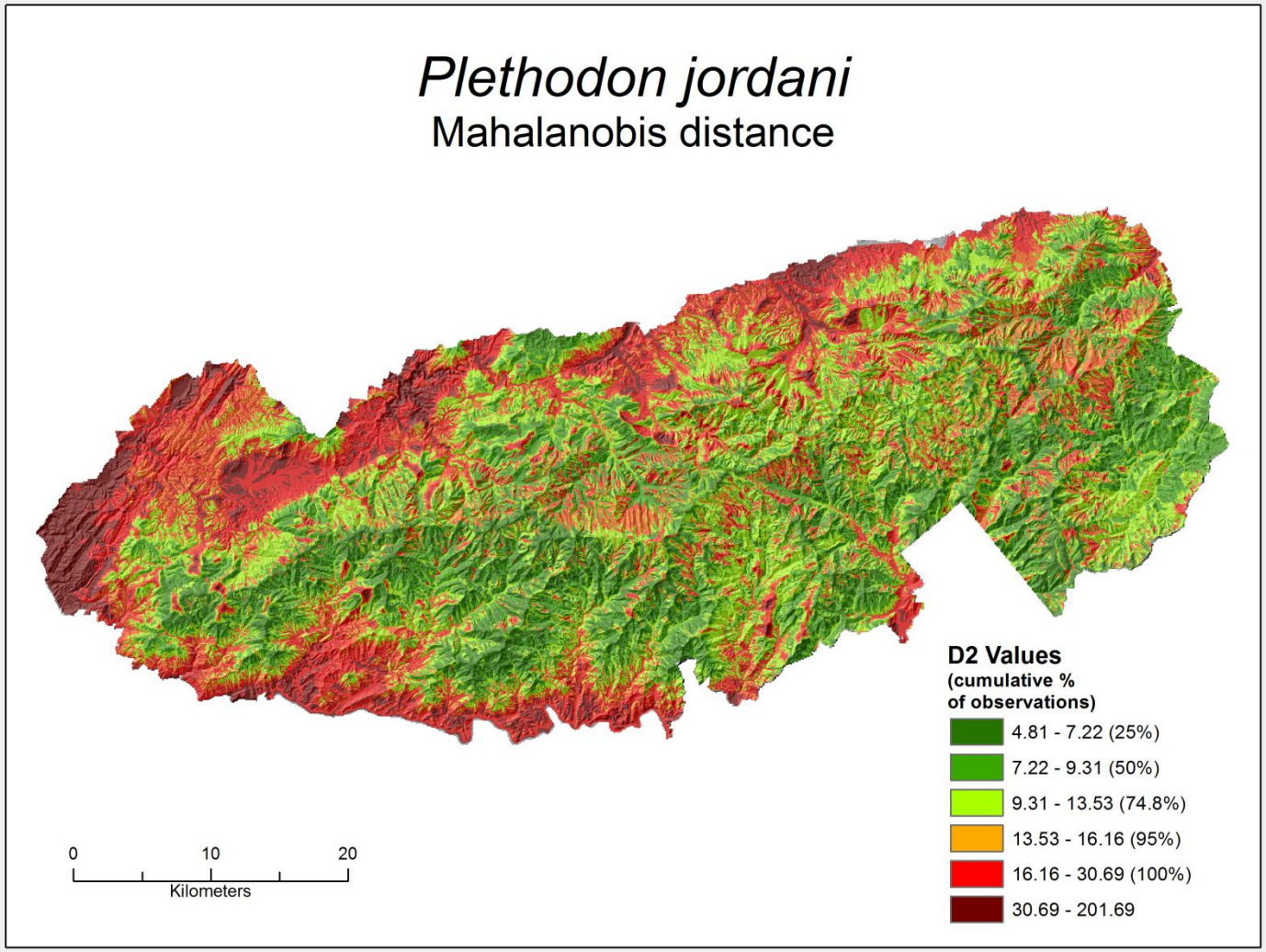

Figure 27. Mahalanobis distance model for *Plethodon jordani.*

Table 19. Coefficients of variation for favorable habitat compared to GSMNP as a whole for each variable used in the *P. jordani* Mahalanobis distance model.

Variable	Favorable	GSMNP
Elevation	0.21	0.35
Historically undisturbed	1.05	1.83
Breakneck/Oconaluftee soil	1.20	2.50
Ditney/Soco soil	1.80	1.04
Herbaceous understory	0.91	1.06
Spruce understory	1.93	5.00
Northern hardwood overstory	1.24	2.00
Distance to streams	0.67	0.80

to streams in favorable habitat ($\bar{x}$ = 195.2, SD = 135.1) and unfavorable habitat ($\bar{x}$ = 184.9, SD = 173.3). However, elevation had a strong positive correlation with distance to streams ($\rho = 0.447$, $p < 0.01$) indicating salamanders were farther from streams as elevation increased.

The remaining categorical variables were historically undisturbed areas, Breakneck/Oconaluftee soils, Ditney/Soco soils, herbaceous understory, spruce understory, and northern hardwood overstory. Historically undisturbed areas comprised 27.3% of favorable habitat and 16.4% of unfavorable areas. Salamanders were more likely to be in undisturbed areas at higher elevations ($r_{pb} = 0.28$, $p < 0.01$) and in areas of spruce understory ($\phi = 0.21$, $p < 0.02$). Locations in Breakneck-Oconaluftee soils were more probable with increasing elevations ($r_{pb} = 0.52$, $p < 0.01$). The soils were also farther from streams ($r_{pb} = 0.32$, $p < 0.01$), and comprised 17.4% of favorable habitat versus 10.2% of unfavorable areas. At lower elevations, the salamander was more likely to be in Ditney-Soco soils ($r_{pb} = -0.49$, $p < 0.01$) which were a dominant soil in 56.9% of favorable habitat versus 35.6% of unfavorable areas. Only 2.1% of the park was spruce understory, but 6.3% of predicted favorable habitat was comprised of this vegetation. Spruce understory was also strongly correlated with elevation ($r_{pb} = 0.39$, $p < 0.01$). Herbaceous understory was the only variable not correlated with elevation or any other non-similar habitat variable. Herbaceous understory comprised 52.2% of favorable habitat and 39.1% of unfavorable areas. Lastly, northern hardwood switched in this model to being less proportional in favorable habitat than unfavorable areas, which would have made it a candidate for removal in the first step.

3.4. Model Comparison

I compared the classification rate of each model using the true presence and absence data from each species and extrapolating the results from each map. The rates from the each case-control method presented earlier were slightly different than the case-control percentages presented here because of small shifts in the grid systems between GIS layers. The overall pattern remains the same.

Desmognathus conanti had the best total classification rate in the case-control model and worst in the use-availability model (Table 20). However, use-availability was more consistent between presences and absences than case-control or D^2 which had a larger dichotomy between presence and absence prediction accuracy.

Desmognathus ocoee models all had similar total classification rates (Table 21). However, the use-availability model predicted less than 50% of presences correctly. Case-control and D^2 both had high presence and absence rates.

Plethodon jordani classification rates were similar to *D. ocooe* rates with similar total classification rates and a low presence rate for the use-availability model (Table 22). D^2 model were better at predicting presence than the case-control model, but the case-control did better at predicting absences than the D^2 model.

Table 20. Classification rates of the *D. conanti* models using the true presence-absence data from the case-control model.

Model	% presence correct	% absence correct	Total % correct
Case-control	92.7	48.7	71.2
Use-availability	75.0	60.5	67.9
Mahalanobis distance	91.1	48.7	70.4

Table 21. Classification rates of the *D. ocoee* models using the true presence-absence data from the case-control model.

Model	% presence correct	% absence correct	Total % correct
Case-control	79.5	89.0	85.7
Use-availability	48.7	98.6	81.3
Mahalanobis distance	92.3	78.1	83.0

Table 22. Classification rates of the *P. jordani* models using the true presence-absence data from the case-control model.

Model	% presence correct	% absence correct	Total % correct
Case-control	74.8	89.1	82.4
Use-availability	64.1	91.8	78.8
Mahalanobis distance	97.7	66.0	80.9

CHAPTER FOUR

4 DISCUSSION

4.1 THE ROLE OF ELEVATION IN HABITAT MODELING

Elevation was the key determinant for habitat across all models for all three species. However, elevation is likely a proxy for an unmeasured variable or combination of variables. As lungless and terrestrial amphibians, plethodontid salamanders need moist environments for respiration and to stay hydrated (Petranka 1998) but elevation does not fill that need. More direct measurements such as precipitation and temperature often have close relationships with elevation (Barry 2008), yet are not available at appropriate scales. Elevation may also reflect the combination of variables such as vegetation and geology that would be difficult to combine without precise knowledge. In other circumstances, an association with elevation may be a result of competition with another species. One species may restrict another to a range of elevation that is not reflective of where the restricted species may exist without the competition (Hairston 1981). Without measurements of those "missing" variables, one can only speculate as to the true meaning of an elevation variable for each species.

4.2 *DESMOGNATHUS CONANTI*

4.2.1 MODELS OF DISTRIBUTION

The spotted dusky salamander is a generalist species that exists at the widest geographical range of the three species. Within GSMNP, the species is known to exist at lower portions of the park (Dodd 2004), which is not only the majority of the western side

but also the stream-carved valleys surrounded by higher elevations. All three models supported the existing knowledge of distribution with regard to elevation range in the Smokies. In fact, elevation was the only variable to exist across all three model types. Each of those models also indicated elevation was the major contributor for distribution of the species within GSMNP. The negative association with elevation is reflective of the distribution of *D. conanti* in low elevations of the southeastern United States (Petranka 1998).

Additionally, each model also used historical disturbance as a predictor of distribution but in different ways. The Mahalanobis distance and use-availability models both included historically settled areas as predictors of presence. The case-control model used historically undisturbed areas as a predictor of absence of the salamander. Either way, the models indicated favorable habitat was more likely in settled areas than undisturbed areas. Whereas settled areas are largely associated with lower elevations, the association with other disturbance at low elevations was not apparent. Settled areas are usually in close proximity to streams, indicating the variable may be a proxy for soil moisture. These areas are probably still going through vegetative succession (Hyde and Simons 2001), and the impact of further succession on the species would be interesting.

All the other variables only existed in one model and were not shared between models. Of those remaining variables, several included in the Mahalanobis distance model were important in delineating the distribution of favorable habitat. Proximity to streams was perhaps the most important of these and showed a very strong correlation with elevation. The strength of that relationship was interesting. At low elevations, distance from streams was the deciding factor as the only areas of unfavorable habitat were those

that were relatively far away from any stream. Alternatively, at higher elevations, elevation was the deciding factor because a close proximity to streams did not matter anymore for good habitat after about 1,100 meters. The good habitat even appeared to be farther away from streams as elevation increased. I am unclear whether this indicates an adequate amount of moisture in the soil at higher elevations or if the pattern shows an avoidance of streams that become too cold for the species.

Other contributing habitat variables were the soil types. Ditney/Soco soils are both metasandstone units (quartz and sand fused together) and have a tendency to form with large rock outcrops. These rock outcrops are favorable to salamanders as the crevices provide cool, shaded, moist micro-habitats as well as a place to hide (Green and Pauley 1987, Pauley 1998, Petranka 1998). The salamanders also exist in Junaluska soils which are mostly siltstone, but have colluviums of metasandstone originating from Ditney/Soco soil units. The relationship with Ditney/Soco soils was stronger in higher elevations and the association with Junaluska soils was stronger in the lower elevations of the species' range. Other studies in the eastern United States have also shown an association with plethodontid habitat and sandstone geology (Dillard 2007).

Remaining habitat characteristics such as overstory vegetation appeared less important. The strength of variables such as floodplain forests in the use-availability logistic regression model may be largely tied to the fact that they are next to streams and rivers which result in moisture for respiration and colluvium for shelter. Oak-hickory forests also appeared in the Mahalanobis distance model and were associated with Ditney/Soco and Junaluska soils. This is not to say that vegetation was unimportant, but

these models may not have been at the scale where vegetation was a factor (Mitchell et al. 2001).

4.2.2 Model Comparison

The three models for spotted dusky salamander used somewhat different habitat variables with elevation as the key variable of all three. Despite some variation in the use of variables, the extrapolation of the model on the landscape of GSMNP created similar visualizations. The logistic regression models had fewer variables and may be more easily understood by land managers than the 10-variable Mahalanobis distance model.

The logistic regression models could be compared directly using their measures of discrimination and calibration. Of the two types, the use-availability model for *D. conanti* appeared to perform the best. The use-availability model had a better classification rate, a higher pseudo-R^2, and less spread and bias error than the case-control model. The case-control model appeared to place too much negative emphasis on historically undisturbed areas where probability of occurrence dropped severely compared to surrounding habitat. Also, the good habitat on the case-control map appeared too contiguous in lower elevations, which was probably not a realistic representation. I believe the problems with the case-control model were a result of the identification method of *D. conanti*. As with many salamanders, identification can be tricky, and the spotted dusky salamander was not immune from the confusion. During salamander identification, Dodd (2004) considered all lower to mid-elevation dusky salamanders to be *D. conanti*, which meant that no high elevation sites were considered observations for the species. Therefore, the absences I generated using survey sites in the case-control model had a higher concentration in higher

elevations with less historical disturbance. I believe the absences in the use-availability model reduced the impact of potential misidentification by not being biased toward any one variable. The use-availability model had improved performance because it was able to pick up variables based on their use and availability over the whole study area.

The D^2 model appeared to make a more realistic distinction between good and bad habitat than the case-control. However, the appearance of a distinction does not make a model better, and D^2 had a classification rate for both presences and absences nearly identical to the case-control model. The use-availability model had a map more similar to D^2 but had a lower presence and total classification rate. For future studies of *D. conanti*, based on the confusing identification of the species, I suggest using a mixed approach rather than depending on any one model. If one model is necessary, I recommend the use-availability model because it is more balanced between presence and absence predictions.

4.3 *DESMOGNATHUS OCOEE*

4.3.1 MODELS OF DISTRIBUTION

Within GSMNP, the Ocoee salamander exists at the most limited geographical range of the three species. The Ocoee salamander only exists at the highest elevations and not much else is known about its habitat (Dodd 2004). All three models corroborate the literature in that high elevation was a strong variable in GSMNP. Once again, elevation was the only variable to appear in all three models, and it was the lone variable in the case-control logistic regression model. The salamander had a positive association with increasing elevation which was interesting considering the Ocoee salamander is reported to exist in a wide range of elevations outside GSMNP (Petranka 1998). The difference in

elevation range between GSMNP and areas outside the park led me to believe that this was either another case of confusing identification or there was something particularly unique about habitat GSMNP. Without the means to identify these salamanders or analyze GSMNP habitat versus habitat outside the park boundaries, I can only surmise at the difference. Some literature suggests *D. ocoee* is prohibited from lower elevations by *D. imitator* through competitive exclusion over areas closer to streams (Bernardo 2000). Also, the model was reflective of the literature in that the species was more terrestrial in the highest elevations of its range (Petranka 1998), perhaps representative of adequate moisture in areas further from streams.

Most other variables in the models were more troublesome to explain. The use-availability logistic regression and Mahalanobis distance models both used spruce understory but appear to have disagreement in their association. The use-availability model indicated a negative association with spruce understory whereas spruce understory was in the D^2 model acting as a predictor of presence. Interestingly, spruce understory was positively associated with *D. ocoee* presence in a univariate use-availability model. This suggested that the use-availability model had an overwhelming amount of variation explained by elevation and some remaining variation was explained by a negative association with observations, paring the probability estimates a bit. Spruce understory may have been a variable that was picked up merely because its association with high elevation vegetation which coincides with the high elevation *D. ocoee*. Similar conclusions could be made about the other variables in the Mahalanobis distance model. Differently from the D^2 model for *D. conanti*, almost all the variables in the *D. ocoee* models had correlations only with elevation. The variables were probably just a reflection of the

salamander's strong association with elevation and not necessarily explanatory of the species distribution in GSMNP.

4.3.2 Model Comparison

The three models defined good habitat more and more broadly from use-availability to case-control to Mahalanobis distance. In this case, the most parsimonious model was probably the best model of the three. The analysis of Mahalanobis distance combined with the comparison of the logistic regression models led me to be undecided between D^2 and case-control for the best model.

The use-availability logistic regression model included elevation and spruce understory and had an AIC score much higher than a use-availability model with only elevation. However, the case-control set of models only considered elevation in a univariate model to have any predictive power. While the use-availability model had a higher classification rate than the case-control model within each method, the case-control did a much better model did much better job when comparing the methods using the true presence-absence data from the case control. Furthermore, the use-availability did a particularly bad job of predicting presences at only 48.7% of true absences predicted correctly. Combined with a higher AUC and a higher pseudo-R^2, the case-control model seemed to perform the better.

Compared to the case-control model, the total classification rate for D^2 was not much different between the methods. The D^2 total rate was slightly smaller because of a lower rate of predicting absence correctly. That lower rate of absence predictions may indicate that the model defined habitat too broadly across GSMNP. Many variables were

included simply because of a strong association with elevation rather than a strong predictor of presence, and vegetation types such as northern hardwood overstory were given low distance values in areas that were probably beyond the range of the Ocoee salamander. The case-control model restricted the range more without hurting the classification rate much, suggesting those other variables were unimportant. Conversely, the case-control model had a lower presence rate than D^2, indicating a definition of habitat that was slightly too restrictive, probably due to its simplicity with one variable. The Ocoee salamander's unique habitat inside GSMNP compared to the areas outside the park was probably reflective of interactions with other species rather than more static habitat covariates. Because of those interactions, this species probably defied the assumption that it was spread optimally in the environment, which contributed to a less than desirable explanation of habitat through modeling.

4.4 *PLETHODON JORDANI*

4.4.1 MODELS OF DISTRIBUTION

The red-cheeked salamander was the species at a geographic medium in GSMNP between the wide-spread spotted dusky salamander and the limited Ocoee salamander. This was reflected in the models with the models providing varying degrees of geographic limitation. Out of the three species, the three models for *P. jordani* had the most visibly different outputs on the landscape. As reflected across all three species, elevation was again a major factor in determining red-cheeked salamander habitat. Each model placed a different degree of importance on elevation, though. The use-availability logistic regression model was the most restrictive with good habitat above 1300 m, while the Mahalanobis

distance model had good habitat in areas as low as 700 m. Being that *P. jordani* is endemic to GSMNP (Dodd 2004), elevations of 700 m are probably too low and would allow for the salamander to exist outside the boundaries of the park. Two models, case-control and D^2, also indicated that distance to stream mattered. In both models, the salamander was relatively closer to streams, but the D^2 model also displayed a similarity to the two other plethodontids in that *P. jordani* became more terrestrial as elevation increased.

The rest of the variables in the three models suffered from many of the problems of the *D. ocoee* models. Many variables seemed to be strongly correlated with elevation only. In the D^2 model, Breakneck/Oconaluftee soils were relatively strong predictors of habitat. Several researchers have documented the relationship of sandstone and emergent rock as beneficial to woodland salamander habitat (Green and Pauley 1987; Pauley 1998; Dillard 2007), but it is difficult to separate the soils from elevation as they have a strong correlation with elevation. The same can be said for undisturbed areas which generally occur at high elevations. However, areas exist on the eastern side of the park that are high elevation with much disturbance and red-cheeked salamander habitat was not predicted in those locations. Similar to the Ocoee salamander, spruce understory was a predictor of presence in the D^2 model but a predictor of absence in the case-control logistic regression model. Again, spruce was also positively associated in a univariate case-control logistic regression model. The mixed results of this variable between models, suggested a strong association with elevation but not necessarily a good predictor of habitat. Additionally, pine understory appeared in the case-control logistic regression model as a predictor of absence and was not in the D^2 model as a predictor of presence. Pine is only at lower elevations and

primarily served to better predict absence in the case-control model probably because it is associated with drier, more historically disturbed habitats.

4.4.2 Model Comparison

The three models for *P. jordani* shared similarities between the models for *D. ocoee*. This was apparent in the fact that they are both relatively high elevation species. In fact, the use-availability models for *P. jordani* and *D. ocoee* used the same two variables and the output was an almost identical map. Similarly, because of the strong association with elevation, the models for *P. jordani*, especially D^2, may have included other variables that were not important. The D^2 model overestimated red-cheeked salamander habitat (predicted a wider geographical range) by including too many unimportant variables whearas the use-availability logistic regression model underestimated the range, and restricted habitat, by placing too much weight on elevation. The case-control model was an acceptable medium between the other two models.

The case-control seemed to perform the best of the three models. It included only four variables versus eight in D^2 which combined with a more balanced classification rate of presences and absences, gave me preference for the case-control model. Even when compared to a more parsimonious use-availability model, the case-control model had a higher overall classification rate (82.4% vs. 78.8%), a much higher classification rate of presences (74.8% vs. 64.1%), a higher AUC (0.92 vs. 0.83), and a higher pseudo-R^2 (0.65 vs. 0.41). The case-control model also had a very little spread error and bias. While the model could to better in an explanatory fashion to distinguish between variables correlated with

high elevation, the case-control model does a sufficient job for a set of data not collected for modeling purposes.

CHAPTER FIVE

5 CONCLUSIONS

The overarching goal of my study was to create habitat models of three plethodontid salamander species (*Desmognathus conanti*, *D. ocoee*, and *Plethodon jordani*) in GSMNP. As literature is lacking or inconclusive on salamander habitat at a landscape-scale, the process was largely exploratory. To investigate the relationships between salamanders and their habitats, I used three separate models (two logistic regression models and one Mahalanobis distance) for each species to gain a more robust view of the relationships. The secondary objective was to compare the different modeling methods within and across the three species. I did this using both objective methods based on quantitative metrics as well as subjective methods based on weighing the benefits and pitfalls of choosing one model over another. The results of this study are unique in that no other study has analyzed the distributions of these species across GSMNP.

Interestingly, the limited qualitative information about habitat for the three species was often unimportant for determining habitat. Many variables, such as northern aspects, that were listed as good habitat in literature and appeared common in my preliminary statistics turned out to be nothing more than salamanders using habitat as it was available. The most important variable was elevation, and it was overwhelmingly the limiting factor in all three models of all three species. While elevation is probably a proxy for other variables such as precipitation or temperature, its strength in the models was clear. Most other variables in the models were secondary to elevation. *Desmognathus conanti*, as expected, is a low elevation species that may also have a preference for habitats that have

been somewhat disturbed in the past. *Desmognathus ocoee* appeared to have the strongest relationship with elevation and was only predicted in the highest elevations of GSMNP. Other variables seemed to be much less important than elevation. Finally, *Plethodon jordani* was also a high elevation species, but was not as restricted as *D. ocoee*. A closer proximity to streams seemed more important to *P. jordani* than to the other two species, although all three were farther away from streams as elevation increased.

All three modeling methods had their benefits and pitfalls that were represented at one time or another in the three species. I found some clear distinctions in which one model severely outperformed or underperformed the other two, but making a definitive decision on which method was outright better was unwarranted. The best method to use was not only dependent on the rigors of that method, but also on the reliability of the fieldwork and the availability of variables to include in the model. Each had a case at some point in the process.

Logistic regression models had the obvious problem of defining absence. Of the two methods, use-availability with randomly generated absences seemed to work better only in the *D. conanti* model. This was probably because of the decision to lump all dusky salamander suspects in low elevations as *D. conanti*, which resulted in no presence locations above a predefined boundary. The case-control logistic regression models performed better in the other two species, where identification was not as much of a problem. This made sense because absences determined by methods in the field should be more reliable than absences determined by randomly generating locations that were likely, on occasion, to fall within good habitat. Overall, logistic regression methods tended to create a more parsimonious model that was easy to understand from the mathematical

perspective. Also, models and variables in logistic regression can be accepted or rejected based on significance. The decision on an acceptable significance level may be subjective, but the measure of probability that is a p-value is mathematically objective. However, logistic regression also has a tendency to eliminate variables that may be important habitat characteristics in favor of an overpowering variable. Logistic regression also does not consider variable interactions unless the modeler specifically enters an interaction term which, without literature on habitat, requires much guesswork and an exponentially additional modeling timeframe for every combination considered.

Mahalanobis distance takes advantage of the interaction between variables that is often missing from logistic regression. The downside is analyzing variables and their effect on the output distribution. Significance plays little role in determining good and bad models and variables. Which variables are most important, the strength of the interaction between two variables, or determining if a habitat combination of good habitat is any different than available combinations are all problems without clear solutions in the literature. Also, by using the mean and a variance-covariance matrix, the method is meant for using continuous variables that exist in a gradient over the landscape. Those continuous variables are naturally going to have correlation that is strong for some and weak for other covariates. But categorical variables are binary and treat all observations in the reference category equally. Mathematically, what does the mean of several 1's and 0's represent? Ecologically, should a variable with several recorded presences be detrimental to the species?

Each categorical variable included in the D^2 model is a variable with several observed presences, and while the variable may not be particularly good, it is not likely to

be bad for the individual. For binary variables in which observations were recorded, if less than 50% of the observations were in that category such as Spivey soil, then the mean will be less than 0.5. With a mean of less than 0.5, it is possible that a location within Spivey soil will have a larger D^2 value than a location in another soil with all other variables the same. Despite the fact that salamanders were observed in Spivey soil, now it has effectively become a predictor of absence within the model. Often, this problem may be negated by strong relationships with Spivey soil and other variables, but the problem could still, and may often, occur. Dealing with these issues makes Mahalanobis distance troublesome, especially for an exploratory model where ecological knowledge is not so clear.

These three models were largely exploratory because of little literature on habitat requirements for salamanders at a scale coarser than the individual. In that sense, none of these methods can be outright accepted or eliminated, but the case-control logistic regression seems to be the most reliable of the three. Mahalanobis distance did seem to work best for *D. conanti*, but that may be a result of the uncertain identification of the spotted dusky salamander. In the two other species, the case-control logistic models provided a high classification rate for both presences and absences and a more intuitive answer for what determines the species habitat. Mahalanobis distance models have trouble distinguishing between what is actually explanatory and what is simply associated with another strong variable. Independent fieldwork meant for model creation would greatly improve the models (Fortin and Dale 2005). Fieldwork for the purpose of parsing some of the variables from their inherent correlations would provide more insights into what variables are ultimately limiting to salamander habitat.

LITERATURE CITED

Aldrich, J.H. and F.D. Nelson. 1982. Linear probability, logit, and probit models. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-045. Beverly Hills, USA, and London, UK.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267–281 *in* B.N. Petrov and F. Csaki, eds. Second International Symposium on Information Theory. Akademiai Kiado, Budapest.

Anderson, D.R., K.P. Burnham, and W.L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912–923.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101–118.

Barry, R.G. 2008. Mountain weather and climate. Cambridge University Press, Cambridge, United Kingdom.

Beers, T.W., P.E. Dress, and L.C. Wensel. 1966. Aspect transformation in site productivity research. Journal of Forestry 64:691–692.

Bentley, J.M., C.P. Catterall, and G.C. Smith. 2000. Effects of fragmentation of Araucarian vine forest on small mammal communities. Conservation Biology 14:1075–1087.

Bernardo, J. 2000. Early life histories of dusky salamanders, *Desmognathus imitator* and *D. wrighti*, in a headwater seepage in Great Smoky Mountains National Park, USA. Amphibia-Reptilia 21:403–407.

Bewick, V., L. Cheek, and J. Ball. 2005. Statistics review 14: logistic regression. Critical Care 9:112–118.

Boestch, J., F.T. van Manen, and J.D. Clark. 2003. Predicting rare plant occurrence in Great Smoky Mountains National Park. Natural Areas Journal 23:229–237.

Brotons, L., W. Thuiller, M.B. Araújo, and A.H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27:437–448.

Browning, D.M., S.J. Beaupré, and L. Duncan. 2005. Using partitioned Mahalanobis D^2 (k) to formulate a GIS-based model of timber rattlesnake hibernacula. Journal of Wildlife Management 69:33–44.

Boyce, M.S., P.R. Vernier, S.E. Nielsen, and F.K.A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling 157:281–300.

Burnham, K.P. and D.R. Anderson. 1998. Model selection and multimodel inference: a

practical information-theoretic approach. Springer, New York, NY.

Center for Remote Sensing and Mapping Science. 2004. Digital vegetation maps for the Great Smoky Mountains National Park. Department of Geography, University of Georgia, Athens, GA.

Clark, J.D., J.E. Dunn, and K.G. Smith. 1993. A multivariate model of female black bear habitat use for a geographic information system. Journal of Wildlife Management 57:519–526.

Crawford, J.A. and R.D. Semlitsch. 2008a. Post-disturbance effects of even-aged timber harvest on stream salamanders in southern Appalachian forests. Animal Conservation 11:369–376.

Crawford, J.A. and R.D. Semlitsch. 2008b. Abiotic factors influencing abundance and microhabitat use of stream salamanders in southern Appalachian forests. Forest and Ecology Management 255:1841–1847.

Crespi, E.J., L.J. Rissler, and R.A. Browne. 2003. Testing Pleistocene refugia theory: phylogeographic analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. Molecular Ecology 12:969–984.

Davic, R.D. and H.H. Welsh. 2004. On the ecological roles of salamanders. Annual Review of Ecology, Evolution, and Systematics 35:405–434.

Delcourt, H.R., and P.A. Delcourt. 1997. Pre-Columbian Native American use of fire on southern Appalachian landscapes. Conservation Biology 11:1010–1014.

Dettmers, R., D.A. Buehler, and J.G. Bartlett. 2002. A test and comparison of wildlife habitat modeling techniques for predicting bird occurrence at a regional scale. Pages 607–615 *in* J.M. Scott, P.J. Heglund, and M.L. Morrison, eds. Predicting species occurrences: issues of scale and accuracy. Island Press, Covelo, CA.

Dillard, L.O. 2007. Distribution and habitat of the threatened Cheat Mountain salamander (*Plethodon nettingi*) at multiple spatial scales in West Virginia forests. MS Thesis, University of Wisconsin, Stevens Point, WI.

Dillard, L.O., K.R. Russell, and W.M. Ford. 2008. Macrohabitat models of occurrence for the threatened Cheat Mountain salamander, *Plethodon netting*. Applied Herpetology 5:201–224.

Diller, L.V. and R.L. Wallace. 1996. Distribution and habitat of *Rhyacotriton variegates* in managed, young growth forests in north coastal California. Journal of Herpetology 30:184–191.

Dodd, C.K. 2004. Amphibians of Great Smoky Mountains National Park. University of Tennessee Press, Knoxville, TN.

Dye, A.S. 2003. Developing habitat suitability models: an example from Great Smoky Mountains National Park, Tennessee and North Carolina, USA, using the land snail *Vitrinizonites latissimus* Lewis. MS Thesis, University of Tennessee, Knoxville, TN.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41:263–274.

Etherington, T.R., A.I. Ward, G.C. Smith, S. Pietravalle, and G.J. Wilson. 2009. Using the Mahalanobis distance statistic with unplanned presence-only survey data for biogeographical models of species distribution and abundance: a case study of badger setts. Journal of Biogeography 36:845–853.

Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Feng, T., F.T. van Manen, N. Zhao, and F. Wei. 2009. Habitat assessment for giant pandas in the Qinling Mountain region of China. Journal of Wildlife Management 73:852–858.

Ford, W. M., B. R. Chapman, M. A. Menzel, and R. H. Odom. 2002. Stand age and habitat influences on salamanders in Appalachian cove hardwood forests. Forest Ecology and Management 155:131–141.

Fortin, M.J. and M.R.T. Dale. 2005. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, United Kingdom.

Funk, V.A. and K.S. Richardson. 2002. Systematic data in biodiversity studies: use it or lose it. Systematic Biology 51:303–316.

Gauch, H.G. and G.B. Chase. 1974. Fitting the Gaussian curve to ecological data. Ecology 55:1377–1381.

Gavashelishvili, A. 2004. Habitat selection by East Caucasian tur (*Capra cylindricornis*). Biological Conservation 120:391–398.

Glenn, E.M., M.C. Hansen, and R.G. Anthony. 2004. Spotted owl home-range and habitat use in young forests of western Oregon. Journal of Wildlife Management 68:33–50.

Green, N.B. and T.K. Pauley. 1987. Amphibians and reptiles in West Virginia. University of Pittsburgh Press, Pittsburgh, PA.

Griffin, S.C., M.L. Taper, R. Hoffman, L.S. Mills. 2010. Ranking Mahalanobis distance models

for predictions of occupancy from presence-only data. Journal of Wildlife Management 74:1112–1121.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N.G. Yoccoz, A. Lehmann, and N.E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. Conservation Biology 20:501–511.

Hairston, N.G. 1949. The local distribution and ecology of the plethodontid salamanders of the southern Appalachians. Ecological Monographs 19:47–73.

Harrod, J., P.S. White, and M.E. Harmon. 1998. Changes in xeric forests in western Great Smoky Mountains National Park, 1936–1995. Castanea 63:346–360.

Hirzel, A.H., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling 145:111–121.

Highton, R. 1995. Speciation in eastern North American salamanders of the genus *Plethodon*. Annual Review of Ecology and Systematics 26:579–600.

Hosmer, D.W. and S. Lemeshow. 1989. Applied Logistic Regression. John Wiley & Sons, New York, NY.

Houk, R. 1993. Great Smoky Mountains National Park. Houghton Mifflin Company, New York, NY.

Hyde, E.J. and T.R. Simons. 2001. Sampling plethodontid salamanders: sources of variability. Journal of Wildlife Management 65:624–632.

Johnson, C.J. and M.P. Gillingham. 2005. An evaluation of mapped species distribution models used for conservation planning. Environmental Conservation 32:117–128.

Johnson, C.J., S.E. Nielsen, E.H. Merrill, T.L. McDonald, and M.S. Boyce. 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. Journal of Wildlife Management 70:347–357.

Karpanty, S.M., J.D. Fraser, J. Berkson, L.J. Niles, A. Dey, E.P. Smith. 2006. Horseshoe crab eggs determine red knot distribution in Delaware Bay. Journal of Wildlife Management 70:1704–1710.

Keating, K.A. and S. Cherry. 2004. Use and interpretation of logistic regression in habitat-selection studies. Journal of Wildlife Management 68:774–789.

Keister, A.R., J.M. Scott, B. Csuti, R.F. Noss, B. Butterfield, K. Sahr, and D. White. 1996. Conservation prioritizations using GAP data. Conservation Biology 10:1332–1342.

Knick, S.T. and J.T. Rotenberry. 1998. Limitations to mapping habitat use areas in changing

landscapes using the Mahalanobis distance statistic. Journal of Agricultural, Biological, and Environmental Statistics 3:311–322.

Kozak, K.H. and J.J. Weins. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60:2604–2621.

Kozak, K.H. and J.J. Wiens. 2007. Climatic zonation drives latitudinal variation in speciation mechanisms. Proceedings of the Royal Society B 274:2995–3003.

Levin, S.A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.

Lutolf, M., F. Kienast, and A. Guisan. 2006. The ghost of past species occurrence: improving species distribution models for presence-only data. Journal of Applied Ecology 43:802–815.

Madison, D.M. 1969. Homing behaviour of the red-cheeked salamander, *Plethodon jordani*. Animal Behaviour 17:25–39.

Means, D.B. 2000. Southeastern U.S. coastal plain habitats of the Plethodontidae. The importance of relief, ravines, and seepage. Pages 287–302 *in* R.C. Bruce, R.G. Jaeger, and L.D. Houck, eds. The Biology of Plethodontid Salamanders. Kluwer Academic, New York, NY.

Mitchell, M.S., R.A. Lancia, and J.A. Gerwin. 2001. Using landscape-level data to predict the distribution of birds on a managed forest: effects of scale. Ecological Applications 11:1692–1708.

National Park Service. 2007. Vegetation disturbance history at Great Smoky Mountains National Park, Tennessee and North Carolina. Great Smoky Mountains National Park, Resource Management and Science. Geospatial dataset 1045868.

National Park Service. 2010a. Ranking report for recreation visits in 2010. National Park Service Public Use Statistics Office. Accessed 15 November 2011. http://www.nature.nps.gov/stats/index.cfm

National Park Service. 2010b. Great Smoky Mountains: Natural Features & Ecosystems. Updated 11 August 2010. United States National Park Service. Accessed 15 November 2011. http://www.nps.gov/grsm/naturescience/naturalfeaturesandecosystems.htm

Parker, A.J. 1982. The topographic relative moisture index: an approach to soil-moisture assessment in mountain terrain. Physical Geography 3:160–168.

Pauley, B.A. 1998. The use of emergent rocks as refugia for the Cheat Mountain salamander, *Plethodon nettingi* Green. MS Thesis. Marshall University, Huntingdon, WV.

Pearce, J. and S. Ferrier. 2000. Evaluating the predictive performance of habitat models

developed using logistic regression. Ecological Modelling 133:225–245.

Peng, C.J., K.L. Lee, and G.M Ingersoll. 2002. An introduction to logistic regression analysis and reporting. The Journal of Educational Research 9:3–14.

Pereira, J.M.C. and R.M. Itami. 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering and Remote Sensing 57:1475–1486.

Petranka, J. 1998. Salamanders of the United States and Cananda. Smithsonian Institute Press, Washington, D.C.

Phillips, S.J., M. Dudik, and R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Proceedings of the 21st International Conference on Machine Learning.

Phillips, S.J. and M. Dudik. 2008. Modeling species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

Pyle, C. 1988. The type and extent of anthropogenic vegetation disturbance in the Great Smoky Mountains before National Park Service acquisition. Castanea 53:183–196.

Rao, C.R. 1952. Advanced statistical methods in biometric research. John Wiley and Sons, New York, NY.

Rotenberry, J.T., S.T. Knick, and J.E. Dunn. 2002. A minimalist approach to mapping species' habitat: Pearson's planes of closest fit. Pages 281–289 *in* J.M. Scott, P.J. Heglund, and M.L. Morrison, eds. Predicting species occurrences: issues of scale and accuracy. Island Press, Covelo, CA.

Russell, K.R., T.J. Mabee, and M.B. Cole. 2004. Distribution and habitat of Columbia torrent salamanders at multiple spatial scales in managed forests of northwestern Oregon. Journal of Wildlife Management 68:405–417.

Sarre, S., G.T. Smith, and J.A. Meyers. 1995. Persistence of two species of gecko (*Oedura reticulata* and *Gehyra variegata*) in remnant habitat. Biological Conservation 71:25–33.

Sattler, T., F. Bontadina, A.H. Hirzel, and R. Arlettaz. 2007. Ecological niche modeling of two cryptic bat species calls for a reassessment of their conservation status. Journal of Applied Ecology 44:1188–1199.

Scott, J.M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards, J. Ulliman, and R.G. Wright. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monographs 123:1–41.

Stanely, T.R. and K.P. Burnham. 1998. Information-theoretic model selection and model

averaging for closed-population capture-recapture studies. Biometrical Journal 40:475–494.

Stauffer, D.F. 2002. Linking populations and habitats: Where have we been? Where are we going? Pages 53–61 *in* J.M. Scott, P.J. Heglund, and M.L. Morrison, eds. Predicting species occurrences: issues of scale and accuracy. Island Press, Covelo, CA.

Stebbins, R.C. and N.W. Cohen. 1995. A natural history of amphibians. Princeton University Press, Princeton, NJ.

Stottard, M.A. and J.P. Hayes. 2005. The influence of forest management on headwater stream amphibians at multiple spatial scales. Ecological Applications 15:811–823.

Swihart, R.K., T.M. Gehring, M.B. Kolozsvary, and T.E. Nupp. 2003. Responses of 'resistant' vertebrates to habitat loss and fragmentation: the importance of niche breadth and range boundaries. Diversity and Distributions 9:1–18.

Tarkhnishvili, D., I. Serbinova, and A. Gavashelishvili. 2009. Modelling the range of Syrian spadefoot toad (*Pelobates syriacus*) with combination of GIS-based approaches. Amphibia-Reptilia 30:401–412.

United States Department of Agriculture, Natural Resources Conservation Service. 2008. Soil Survey Geographic (SSURGO) Database for Great Smoky Mountains, Tennessee and North Carolina. Accessed May 2009. http://soildatamart.nrcs.usda.gov.

United States Department of Agriculture, Natural Resources Conservation Service. 2009. Soil survey of Great Smoky Mountains National Park, Tennessee and North Carolina.

Van Manen, F.T., J.A. Young, C.A. Thatcher, W.B. Cass, and C. Ulrey. 2005. Habitat models to assist plant protection efforts in Shenandoah National Park, Virginia, USA. Natural Areas Journal 24:339–350.

Vitt, L.J., J.P. Caldwell, H.M. Wilbur, and D.C. Smith. 1990. Amphibians as harbingers of decay. Bioscience 40:418.

Welsh, H.H. and S. Droege. 2001. A case for using plethodontid salamanders for monitoring biodiversity and ecosystem integrity of North American forests. Conservation Biology 15:558–569.

Whitaker, R.H. 1967. Gradient analysis of vegetation. Biological Reviews 42:207–264.

Wilds, S.P. 1997. Gradient analysis of the distribution of a fungal disease of *Cornus florida* in the southern Appalachians, Tennessee. Journal of Vegetation Science 8:811–818.

Zaniewski, A.E., A. Lehmann, and J.M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157:261–280.

VITA

Matthew Stephen Kookogey was raised in Newnan, Georgia, and is the oldest of three children. His interest in the natural world began very early perhaps exemplified in his first memory—a chipmunk running across the road. Matthew was awarded a full-tuition scholarship to Georgia Southern University where he graduated in 2006 with a B.A. in Biology. Ultimately turned off by the minutia of biological work, after a couple "lost" years, Matthew discovered geography and decided to pursue a Master's degree at the University of Tennessee. His efforts paid off, and he is currently contracted at UT to work on habitat models to help determine the best route, environmentally, for a new Tennessee highway. Outside of work, Matthew enjoys to read, chase his dog, refinish furniture, travel, but most of all, have a good beer with good friends.